DOCUMENT RESUME

ED 094 042 UD 014 357

AUTHOR	Koslin, Sandra; And Others
TITLE	Guidelines for the Evaluation of Desegregation Programs in School Districts.
INSTITUTION	Riverside Research Inst., New York, N.Y.
PUB DATE	Sep 72
NOTE	174p.
AVAILABLE FROM	Riverside Research Institute, 80 West End Avenue, New York, New York 10023 (Price not available)
EDRS PRICE	MF-$0.75 HC-$7.80 PLUS POSTAGE
DESCRIPTORS	Cognitive Tests; Evaluation Criteria; Evaluation Methods; *Guidelines; *Integration Effects; *Program Evaluation; Program Planning; Racial Balance; Research Design; *Research Methodology; School Districts; *School Integration; Statistical Analysis; Student Characteristics; Transfer Programs

ABSTRACT

These guidelines, which summarize the content of desegregation program evaluation studies, are organized into five chapters. The first is an "Introduction and Overview." Chapter 2 specifies desegregation program variables which are important in planning and designing evaluation studies. The structural and student variables that are required in the evaluation process are discussed. One important structural variable is racial balance, new and objective indicators of which are developed and discussed in chapter 3. Chapter 4 discusses currently available effectiveness measures that school districts can use to determine changes in students' achievement and racial attitude occurring as a consequence of desegregation. Chapter 5 discusses five major types of evaluation study designs which can be used to evaluate the effects of desegregation programs. The advantages and disadvantages of each design approach are discussed in detail, and evaluation issues which school districts must bear in mind while planning and implementing desegregation programs are identified. The nature and extent of the information provided by each design approach, the facilities and capabilities required to execute studies belonging to each design type, and important cost factors associated with each design type are also described. (Author/JM)

ED 094042

RIVERSIDE RESEARCH INSTITUTE

RR

80 West End Avenue / New York, New York 10023 / (212) 873-4000

SEPTEMBER 1972

GUIDELINES FOR THE EVALUATION OF DESEGREGATION PROGRAMS IN SCHOOL DISTRICTS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

UD 014357

Principal Contributors to this Volume

Chapters II, IV and V of the guidelines are by Sandra C. Koslin. Chapter III and the Appendix are by Ben Josephson, Jr. and Richard Pargament.

Table of Contents

		Page
I.	Introduction and Overview	1
II.	Defining the Desegregation Program	5
	A. Specifying Structural Desegregation Objectives	5
	1. What is the nature of the desired change in student body composition?	5
	2. How extensive will the desegregation be?	7
	B. Defining Desegregation Processes	7
	C. Describing Student Characteristics and Predesegregation Environments	7
III.	Racial Balance Indicators	12
	A. The Need for Objective Indicators of Racial Balance	13
	B. A Precise Way to Characterize Changes in School Environments Resulting From Desegregation Programs	24
	1. Preliminaries	24
	2. Difficulties with simple indicators of racial balance	26
	3. A new indicator of racial balance	31
	4. Suggestions for the presentation of school and classroom balance information	47
	C. The Role of Racial Balance in the Monitoring and Evaluation of Desegregation Programs	59
	1. Information to be collected from each school	61

Table of Contents (cont'd)

		Page
	2. Information to be collected from each classroom teacher	62
IV.	Cognitive and Noncognitive Effectiveness Measures for Desegregation Programs	64
	A. Cognitive Effectiveness Measures	65
	1. Scope of this discussion	65
	2. Norm referenced tests	66
	3. Criterion referenced tests	70
	4. Conclusion	73
	B. Noncognitive Effectiveness Measures	75
	1. Importance of assessing noncognitive program outcomes	75
	2. Criteria for noncognitive effectiveness measures	76
	3. The People Test: A measure of social distance	77
	4. The Pick-A-Class Test	87
V.	Evaluation Designs	94
	A. Issues Which Cut Across All Designs	97
	1. Selection of independent and dependent variables	97
	2. Independent variables must be limited in number	103
	3. Dependent variables	105
	4. Advance selection of variables and planning of designs	105

Table of Contents (cont'd)

		Page
	5. Sample size	106
	6. Sampling an adequate number of minority students	107
	7. Accurate identification of student characteristics	109
	8. Stability of the desegregation program	110
	9. Time of testing	110
B.	Designs for Desegregation Programs in Which Only Some Students are Program Participants	111
	1. Type I Designs: Postdesegregation comparisons only	112
	2. Type II Designs: Predesegregation and postdesegregation comparisons	116
	3. Type IIA Designs: Predesegregation and postdesegregation comparisons with random student assignment	125
C.	Types of Designs When All Students are Program Participants	127
	1. Type III Designs: Longitudinal comparisons of scores for the same group of students	128
	2. Type IV Designs: Longitudinal comparisons of scores for different groups of students	135
	3. Type V Designs	139
	4. The special case of grade-a-year desegregation	144
	5. Summary Table	145
Appendix	The Development of Racial Balance Indicators Based on the Probability of Occurrence of School and Classroom Distributions	A-1

Chapter I

Introduction and Overview

An important goal of a racial desegregation program involving changes in the distribution of students to instructional units is to produce positive changes in students' attitudes and achievement. To achieve this goal, a school district first must formulate and execute a plan to alter the distribution of students of different races to the schools and classrooms of the district. Second, the district must carry out an evaluation study to determine whether or not the desired structural changes in school or classroom composition have been achieved, and if they have, to determine what changes in the cognitive and noncognitive behavior of students have occurred as a result of the structural alterations. These two steps are related, since the design of the evaluation study is strongly affected by the particular processes chosen by the district to achieve structural desegregation objectives, as well as by the nature and extent of the desired structural changes reflected in the desegregation objectives themselves.

The particular desegregation objectives chosen by a district and the particular desegregation program adopted to implement these objectives will be determined by many factors: the district's assessment of the need for and the benefits to

be gained from desegregation; the social and political values of the district's population; the availability within the district of funds and other resources; and so on. In dealing with these complex and varied factors to formulate an appropriate desegregation program, a district may easily--and understandably--lose sight of the need to plan an evaluation study concurrently with the formulation of the desegregation program itself. The possibility of inadequate evaluation study planning is heightened by the many technical considerations, frequently unfamiliar to local district personnel, which go into the design of an evaluation study.

If a desegregation program is planned and executed without giving adequate consideration to the study which will evaluate the effects of the program on students' development, evaluation of the program may yield ambiguous information or, in some cases, may not yield any useful information at all. Thus, if the evaluation study design does not go hand in hand with desegregation program planning and implementation, a district may find that it has expended both effort and resources to desegregate but is unable to determine what cognitive and noncognitive effects the desegregation program has had on participating students. In the best interests of the children, a school district should seek to avoid this situation. As with any human endeavor, and any educational program, good intentions do not guarantee success. If a

desegregation program fails to meet its objectives, well-designed and well-executed evaluation studies will reveal this fact, and serve to alert the district to the need to take appropriate remedial action.

The purpose of these guidelines is to summarize the content of desegregation program evaluation studies so that districts can plan desegregation programs that will allow cognitive and noncognitive changes in the behavior of participating students to be determined unambiguously. The remainder of the guidelines is organized into four chapters, each dealing with a different element of the evaluation process.

Chapter II specifies desegregation program variables which are important in planning and designing evaluation studies. The structural (e.g., environmental) and student variables that are required in the evaluation process are discussed, and minimum sets of structural and student variables required for evaluation studies are defined.

One important structural variable is racial balance--the evenness with which minority students are distributed to schools and classrooms of a district. New, objective indicators of racial balance are developed and discussed in Chapter III. These racial balance indicators are required to assess the effectiveness of desegregation programs in meeting stated objectives for structural change, as well as to analyze the effects of desegregation programs on students' attitudes and

achievement.

Chapters IV and V deal directly with important issues in the design of the evaluation studies themselves. Chapter IV discusses currently available effectiveness measures that school districts can use to determine changes in students' achievement and racial attitudes occurring as a consequence of desegregation.

Chapter V discusses five major types of evaluation study designs which can be used to evaluate the effects of desegregation programs. The advantages and disadvantages of each design approach are discussed in detail, and evaluation issues which school districts must bear in mind while planning and implementing desegregation programs are identified. The nature and extent of the information provided by each design approach, the facilities and capabilities required to execute studies belonging to each design type, and important cost factors associated with each design type are also described.

The State Education Department is prepared to support districts in the design and implementation of evaluation studies of desegregation programs. The nature of the support which the Department is prepared to provide is described at appropriate places in the guidelines.

Chapter II

Defining the Desegregation Program

The term "desegregation" has been used by school districts to refer to many different processes. Consequently, a statement by a district that it intends to desegregate does not by itself provide precise enough information to enable formulation of a plan for evaluating the desegregation program. Chapter II outlines the program elements which must be specified by the district in order to permit development of an appropriate evaluation plan.

Key program elements which must be enumerated include specific objectives with respect to restructuring the racial and social class composition of educational units, details of how these changes are systematically to be brought about, and accurate descriptions of the predesegregation characteristics of the student body and the school environments which are to be affected by the desegregation program.

A. Specifying Structural Desegregation Objectives

In order to develop an appropriate plan for evaluating program effectiveness, the school district must be able to answer accurately the following questions regarding its desegregation objectives for the structure of educational units.

1. What is the nature of the desired change in student body composition? The types of change most frequently sought

by districts undergoing desegregation are changes in the racial balance, racial heterogeneity and, to a lesser degree, social class balance and social class heterogeneity of schools and classrooms.

Racial balance refers to the evenness with which minority students are distributed to educational units. Is the program designed to alter the racial balance of classes within schools and/or the racial balance of schools within districts? If so, what is the magnitude of the change desired?

Racial heterogeneity refers to the actual proportion of minority students in a class, grade, or school. (Whereas racial balance refers to the evenness or randomness of distribution, heterogeneity refers simply to the proportion of students available to be distributed.) Is the desegregation program designed to change the racial heterogeneity of schools or classrooms? If so, what is the magnitude of the desired change?

Social class balance and social class heterogeneity are analagous to racial balance and racial heterogeneity, but refer to the socioeconomic (SES) characteristics of educational environments rather than to their racial characteristics. Is the alteration of the social class balance or heterogeneity of schools or classrooms an objective of the desegregation program? If so, what is the nature and extent of the change desired?

2. How extensive will the desegregation be?

a. How many grades are to be involved: one grade, several grades, all grades?

b. How many schools are to be involved: one school, several schools, all schools?

c. What proportion of students in the affected grades and schools are to be involved?

B. Defining Desegregation Processes

A wide variety of processes have been employed by school districts to meet their desegregation objectives. Adequate description of any desegregation program therefore requires that the district specify which method(s) will be used to bring about the desired changes in student body composition, and how it (they) will be implemented (see Table II-1 for examples). If more than one method or more than one mode of implementation is to be used, the district must specify which methods and modes will be used for which students.

C. Describing Student Characteristics and Predesegregation Environments

There are two important reasons for objectively describing the characteristics of students and of educational environments prior to desegregation. First, if it is the objective of the desegregation program to alter the racial or social class composition of schools, grades, or classes, an accurate record of

Table II-1

Desegregation Processes: Some Examples

Methods of Implementation	Modes of Implementation: Transportation	Participation	Assignment
Two schools of contrasting racial/SES composition pool their student bodies, with half the grades being assigned to each school (school pairing or Princeton Plan).	walking versus busing	voluntary versus non-voluntary	random versus non-random
Redistricting of all neighborhood school boundaries so that there is a general reassignment of students.			
Closing of an all-minority school and distribution of students to several schools.			
Opening a new school on the boundary between two racially disparate neighborhoods.			
Transfer of a few minority students to another school within the district or to another district as a demonstration project.			
Creation of an "educational park."			
Consolidation of all students of a given grade in the same building.			
Altering classroom assignment policies within previously desegregated schools, e.g., switching from homogeneous to heterogeneous grouping policies.			

these educational environments must be obtained prior to implementation of the program, so that there will be adequate baseline data to permit determination of whether or not the desired changes have occurred.

Second, an accurate description of student and environmental characteristics is needed so that the research design which is developed (see Chapter V) can take into account as many as possible of the variables or factors which might influence the outcomes of the evaluation study.

The minimal information which must be known for each student are his school, grade, class assignment, race, and sex. Data on SES background, IQ, and prior academic achievement are also desirable (but will be useful only if the measures or indices used are comparable for all students). This set of data for each student will enable the district to describe accurately the nature and distribution of the student body by grade and by school.

Using this information about students, districts can characterize classes, grades, and schools with respect to a number of environmental variables. The minimum set of environmental variables which should be indexed (see Chapter III) consists of classroom racial balance, classroom racial heterogeneity, school racial balance, and school racial heterogeneity, for every grade in which the desegregation program will operate.

If possible, the SES balance and/or heterogeneity of classes and schools should also be assessed (this is a definite requirement whenever a change in SES balance or heterogeneity is a specific program objective), as should the level and variability of IQ and achievement. Moreover, since teacher characteristics such as race, amount of experience, and verbal ability may influence the outcomes of desegregation, it is appropriate to index them. To be of utility in the program evaluation, these variables should be assessed in comparable ways for all classes and for all schools in the grade levels affected by desegregation.

The environmental characteristics, or variables, that are indexed will be used in what is essentially a two-stage evaluation process. In the first stage (described in Chapter III), the values of these variables prior to and following desegregation will be compared. Thus during the first stage of the evaluation process, changes in the values of the environmental variables will themselves be analyzed. These changes constitute the effectiveness measures (dependent data) which will determine the degree of success of the program in accomplishing its objectives for structural change.

If analysis of the environmental indices shows that the desired structural changes have occurred, the district may proceed with the second phase of the evaluation process. In the second stage (described in Chapter V), these same environmental

characteristics become the classificatory (independent or control) variables in terms of which the cognitive and non-cognitive effects of the desegregation program are analyzed.*

However, if the analysis of changes in the environmental indices shows that the variables have not been altered in the desired way, there is no rational basis for the district to proceed to assess any cognitive or noncognitive program effects, since the structural changes which were supposed to define the desegregation program have not taken place.

---

* Even if the desired change in student body composition has occurred, there will be instances in which it will not be worthwhile to assess cognitive and noncognitive program effects. Such instances are those where only a handful of minority students are involved in the program and where the program may be categorized as a "demonstration project." An example would be the case where a small number of urban minority students (e.g., 30 or fewer) drawn from several grade levels are bused to an all white school in their own or a neighboring district. Since the program is apt to be permeated by an air of novelty, since the bussed students are apt to be a highly select group and, most importantly, because there are so few minority students involved at any grade level, it will usually not be worth the expenditure of district resources to undertake the work which would be required to assess cognitive and noncognitive outcomes.

Chapter III

Racial Balance Indicators

The minimum set of environmental variables required to describe school environments for the purpose of performing evaluation studies consists of school racial balance, classroom racial balance, school racial heterogeneity and classroom racial heterogeneity (see Chapter II). School and classroom racial heterogeneity--the proportions of minority students in schools and classrooms--are easy to calculate if sufficient information about student distributions has been collected. School and classroom balance--the evenness with which minority students are distributed to schools and classrooms--are much more difficult to characterize objectively. Sections A and B of this chapter deal with the construction of objective indicators of racial balance, and with the presentation of racial balance information to facilitate decision-making. Section C considers practical problems of data collection, data processing and the use of racial balance information to monitor and evaluate desegregation programs.

A. The Need for Objective Indicators of Racial Balance

A school district contemplating a program which will alter the distribution of minority group children in the schools and classrooms of the district--that is, one which will affect the pattern of racial balance--must be able to characterize this distribution in quantitative and objective terms. Without objective indicators, the extent of racial balance within the classrooms and schools of a district, and the changes in racial balance which occur when racial desegregation programs are implemented, become matters of individual judgement, providing (unfortunately) fuel for an extended debate which can reach no constructive conclusion. Approaching problems of racial balance without objective indicators is like examining a business which keeps no financial records: the financial health of the business and its future prospects become anybody's (or, in this analogy, everybody's) guess.

In these guidelines, school and classroom racial balance have been given precise meanings. The definitions of school and classroom balance, which appear on page 19, directly relate racial balance to the extent to which a distribution of students to schools and classrooms is nondiscriminatory, and are consistent with the intent of desegregation legislation. Furthermore, the indicators which follow from these definitions

are entirely descriptive: they characterize student distributions, but say nothing about the consequences to students of school and classroom environments of varying racial balance. Relating cognitive and noncognitive student behavior to different levels of racial balance (and other variables) is an entirely separate matter, and is left to evaluation studies of the types considered in Chapter V.

Objective indicators of racial balance can be of great use to local administrators and decision-makers in the formulation and implementation of desegregation programs. Specifically, such indicators would serve:

- To identify unambiguously instances of racial imbalance which exist within the district;
- To permit alternative desegregation programs to be sensibly compared. The structural effects of different proposed desegregation programs could be assessed by calculating the changes in the values of the racial balance indicators which each alternative program is expected to produce. If the costs of alternative programs were estimated, cost-effectiveness comparisons could also be carried out;
- To permit desegregation programs to be monitored and evaluated. If the values of racial balance indicators were obtained at the inception and the completion of a desegregation program, the extent and

nature of the racial balance achieved as a result of the program could be determined unambiguously. In instances where desegregation programs extended over several years, changes in the values of the racial balance indicators could be used to monitor progress annually or semiannually.

Establishing the need for quantitative, objective indicators of racial balance is considerably easier than generating the indicators themselves. The discussion in this section of the guidelines will show that care must be exercised in the definition of racial balance indicators if real objectivity and a high degree of practical utility are to be attained. Appropriate racial balance indicators must be sensitive only to those properties of student distributions which are under the control of local administrators, and must not be affected by factors such as district student population size and district racial heterogeneity which local administrators cannot control.

The matter of identifying those properties of student distributions to which racial balance indicators <u>should not</u> be sensitive is basic to the development of useful objective indicators. From the viewpoint of developing racial balance indicators, the most important "uncontrollable" differences are these:

- Size of the district student population;
- Racial heterogeneity of the district student population. This factor includes not only the number of student minority groups in the district (black, Puerto Rican, Indian, oriental, etc.) but also the proportion (i.e., the fraction) of students from each minority group at each grade (or age) level in the district student population. The proportion of minority group students at each grade (or age) level may vary widely from grade to grade as well as between districts. In fact, "minority" students may constitute the "majority" at some grade levels in certain districts.
- Year-to-year changes in student population size and racial heterogeneity. Changes will occur because of the migration of families with children into and out of the district, or because of redistricting.

In effect, these factors represent a set of constraints imposed on local administrators. Local administrators must operate within these constraints to achieve racial balance in the district's schools. Therefore, a condition on the following discussion will be that appropriate definitions of racial balance as well as appropriate formulations of racial

balance indicators must be unaffected by these constraints.* The early identification of these constraints will serve to guide the remainder of the discussion in this chapter.

For a district student population of a given size and racial heterogeneity, the process by which students arrive at particular classrooms begins with the enrollment of the students in the various schools of a district. The administrators in each school then assign students enrolled in that school to classrooms or other instructional units. Investigations of racial balance in schools and districts indicate that the appearance of students in classrooms can be accurately represented by such a two-step process, and furthermore that the first step (the enrollment of students in schools) can be represented independently of the second (the assignment of students to classrooms within schools). The fact that students are not assigned directly to classrooms on a district-wide basis suggests that two "kinds" of racial balance must be considered for each district: school balance (characterizing the outcomes of the enrollment of the district's students in different district schools), and classroom balance (characterizing the outcomes of the assignment

---

* This condition should not be interpreted to mean that the constraints cannot or, in the best interests of the children, should not be changed. Rather, the condition reflects the judgement that alterations in the constraints transcend the authority of local administrators, and that in any given school year administrators can only operate within the constraints to achieve the best possible results.

of students in each school to classrooms); and that the indicator for each "kind" of racial balance should be independent of the other.

This last point--that school and classroom balance are independent quantities--requires some amplification. Just as district administrators do not control the racial heterogeneity of a district's student population, so individual school administrators do not control the racial heterogeneity of the student population enrolled in a particular school.* School administrators must "work with what they have." Thus an indicator of classroom balance should not be affected by the properties of the student populations which school administrators assign to classrooms. School balance and classroom balance are significantly different aspects of district racial balance, and each must be considered separately in assessing the status of racial balance in the district as a whole.

School balance and classroom balance can be defined in operational terms as follows:

---

* School racial heterogeneity is affected by any procedure which governs the enrollment of students in schools, i.e., school racial heterogeneity depends on school racial balance. However classroom balance should be independent of school racial heterogeneity just as school balance should be independent of district racial heterogeneity.

- School balance is the extent to which the proportion of students from a particular minority group at each grade level in every school approaches the proportion of students from that minority group at that level in the district as a whole. Note that the concept of school racial balance is applied separately to each minority group; for each minority group, the proportions referred to are calculated with respect to total populations of the school and the district, regardless of how many other minority groups are represented in those total populations;

- Classroom balance is the within-school counterpart of school balance. Classroom balance is the extent to which the proportion of students from a particular group at each grade level in every class approaches the proportion of students from that minority group at each grade level in the school as a whole. Again, the concept of classroom balance is applied separately to each minority group; for each minority group, the proportions referred to are calculated with respect to the total populations of the classroom and the school, regardless of how many other minority groups are represented in those total populations.

It is the purpose of Section B of this chapter to obtain quantitative indicators of school and classroom balance and to propose ways of presenting racial balance data to administrators, school boards and other decision-makers in school

districts. However, the operating definitions which appear above raise several points which require further discussion before proceeding with the development of the indicators.

First, the definitions of both school and classroom balance treat each minority group separately because, although segregation will certainly alter the behavior of minority group students that are subjected to it, there is no evidence that members of different minority groups will react in the same way to a segregated environment. Aggregated racial balance indicators can be developed which determine school and classroom balance for several minority groups combined. However, it would be impossible to tell from such combined indicators whether students from one minority group were more or less evenly distributed than students from another minority group. Furthermore, each minority group will want to know how its members are distributed in the schools and classrooms of the district. Therefore it appears to be advisable to treat students from each minority group separately. In the event that a district has only a few students from a particular minority group (eight Indians, for example), it would probably be best to handle them on an individual basis. The calculated values of the racial balance indicators developed subsequently in this chapter behave erratically when only a few minority group students are being distributed.

Second, the definitions of both school and classroom balance treat each grade separately because, as discussed earlier in this section, district racial heterogeneity may vary widely from grade to grade within a district, and this is considered to be an "uncontrollable" factor to which the values of balance indicators should be insensitive. However, the fact that each grade is treated separately does not mean that balance indicators for each grade cannot be combined in an attempt to summarize in one number the status of school or classroom balance throughout the district.

The problem with such summary numbers lies in their interpretation. A great deal of information is lost in combining balance indicators for several grades, particularly when racial heterogeneity or racial balance or both vary widely from grade to grade. Under such circumstances, using summaries exclusively can have the effect of obscuring the existence of racial imbalance in certain grades or schools of a district, and result, for example, in a poor choice among alternative desegregation plans. Consequently, district-wide summaries should be used with caution, and never to the exclusion of grade-by-grade and school-by-school balance indicators.

Finally, there may appear to be problems in applying the definitions of school or classroom balance if the schools

within a district depart from the traditional classroom-within-grade instructional pattern in which the same students are always grouped together with the same teacher to receive instruction. In practice, the definitions of school and classroom balance--and the balance indicators which follow from them--can be easily adapted to accommodate different instructional patterns such as ungraded instruction and the periodic regrouping of students in different instructional units. If some schools within a district are ungraded, the ages rather than the grade levels of students can be used as a basis for determining balance indicators. If the classroom to which a student is assigned is an administrative unit (or "home room") that has no relationship to the classes in which instruction is received, classroom balance as defined above may be replaced with "instructional unit" balance, and each instructional class* within each classroom hour considered to be a self-contained classroom. This situation would increase somewhat the amount of data required to determine the analog of classroom balance, but would not significantly complicate the computation of the corresponding racial balance indicator which is carried out on a digital computer. If the task of data collection became burdensome, it could be reduced.

---

* Lunch, assembly and homeroom periods would be omitted.

Sampling procedures have been devised from which the degree of "instructional unit" balance in the student population (for different schools or for the whole district) could be inferred.*

* A detailed description of the sampling approach is outside the scope of these guidelines.

B. A Precise Way to Characterize Changes in School Environments Resulting From Desegregation Programs

1. Preliminaries. The principal purpose of this section of the guidelines is to develop valid, quantitative indicators of school and classroom balance that conform to the operational definitions and are insensitive to the "uncontrollable" factors discussed in the previous section of this chapter. These indicators, together with school and classroom racial heterogeneities, form the minimum set of variables used to describe school environments. The development in this section is descriptive rather than rigorous, and employs examples of simple distributions of students to classrooms or schools to illustrate concepts and problems.

The student distributions used as examples in this section will be presented as shown in Figure III-1. Figure III-1 shows a distribution of students from minority group m (blacks, for instance) in the classrooms (eight in this example) of the fifth grade in a particular district school (labeled s). Note that two kinds of numbers appear in the figure: Those within the heavy rectangle; and those at the edges, or margins, of the rectangle. The numbers within the rectangle constitute the distribution of students from

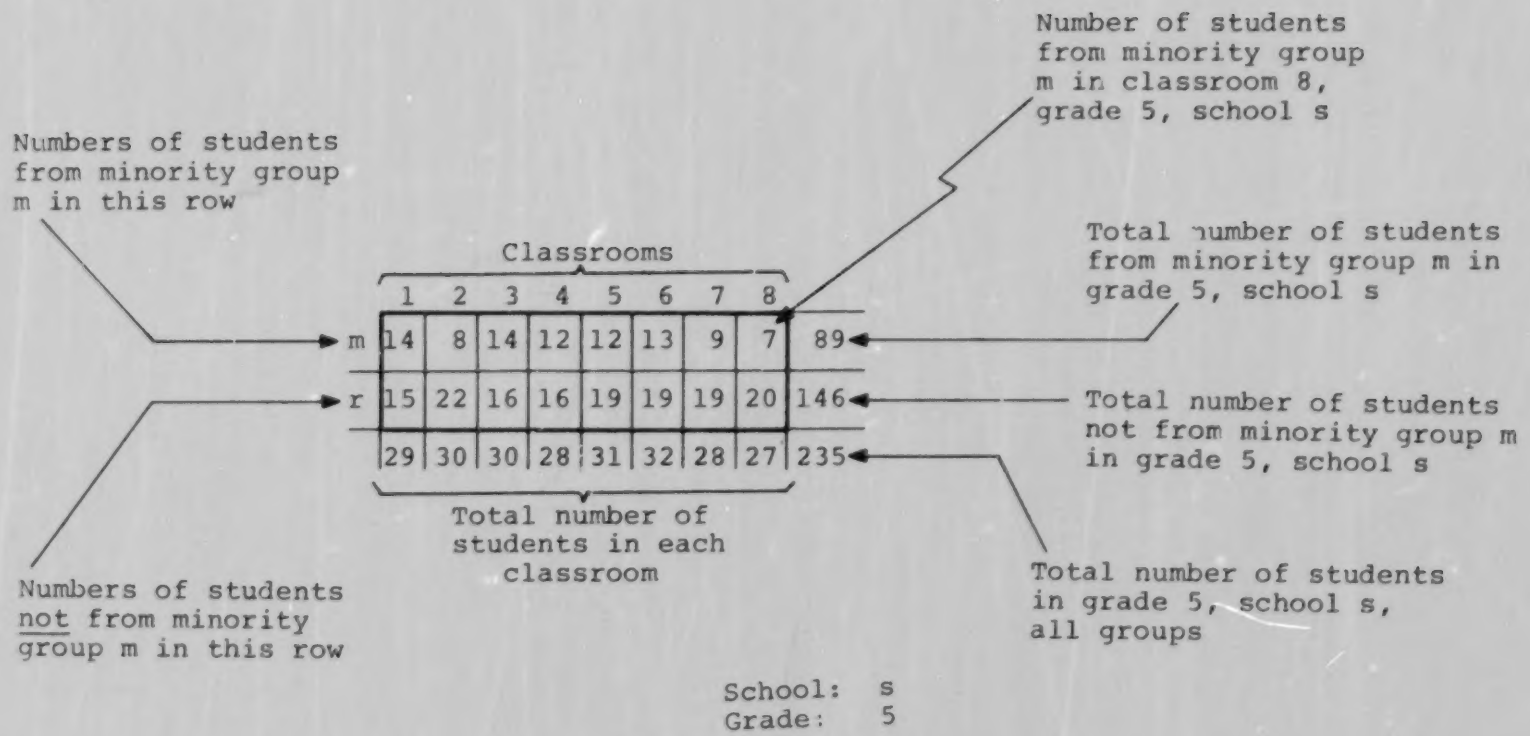

Number of students from minority group m in classroom 8, grade 5, school s
Numbers of students from minority group m in this row
Classrooms
1 2 3 4 5 6 7 8
m 14 8 14 12 12 13 9 7 89
r 15 22 16 16 19 19 19 20 146
29 30 30 28 31 32 28 27 235
Total number of students from minority group m in grade 5, school s
Total number of students not from minority group m in grade 5, school s
Total number of students in each classroom
Numbers of students not from minority group m in this row
Total number of students in grade 5, school s, all groups
School: s
Grade: 5

Fig. III-1. An example of a student distribution. In this example, the distribution of students belonging to a particular minority group to the classrooms of grade 5 in school s is shown in relation to the remaining students in grade 5 and school s.

minority group m (top row), and of students not from minority group m (denoted as r; bottom row), to the eight classrooms of the grade. The marginal numbers are various totals which will be useful at certain stages of the discussion; they are defined directly on the figure.

A similar array can be used to specify the distribution of m-students (from minority group m) and r-students (students from all other groups - the "remainder") in a particular grade to the different schools of a district. In this case, the columns of the array would denote different schools (rather than classrooms). The marginal totals at the right would denote numbers of m-students and r-students in the particular grade in the district, the marginal totals at the bottom would denote the total number of students in the particular grade in each school, and the grand total at the lower right corner would give the total number of students in the district at the particular grade level being examined.

2. Difficulties with simple indicators of racial balance. In order to develop suitable racial balance indicators, a quantitative way must be found to characterize student distributions. The most direct and simplest way to develop an indicator is to work directly with the percentages of minority students (m-students) in each class or school. These percentages can be manipulated and displayed in various ways to

try to describe the extent of school or classroom balance existing in the schools or classrooms of a district.

Indicators constructed from percentages of minority students have been used frequently to measure racial balance. For either the classrooms in a school or the schools in a district, the following quantities can be defined (as usual, for a particular minority group and a particular grade):

- minimum percent. The smallest percentage of minority students found in any classroom (or school);
- maximum percent. The largest percentage of minority students found in any classroom (or school);
- percent range. The difference between the maximum and minimum percents;
- average deviation of percents. Suppose p_1, p_2, p_3... are the percents of minority students in different classes (or schools), p is the percent of minority students in the school (or the district), and n is the number of classes (or schools). Then the average deviation of percents (AD) is defined as

$$AD = \frac{|p_1-p|+|p_2-p|+\ldots+|p_m-p|}{n},$$

where the vertical lines around $|p_i-p|$, indicate that the positive version of the two possible differences,

(p_i-p) and (p-p_i), should be chosen in order to make <u>each term</u> in the AD positive.*

To see why these quantities should not be used as indicators of racial balance, consider the four examples of classroom distributions shown in Figure III-2 for a particular minority group. In the first distribution, classes 1 through 6 are perfectly balanced (the proportion of minority students in these classes is the same as the proportion of minority students at that grade level in the school), while classes 7 and 8 are maximally unbalanced. In the second distribution, all classes are maximally unbalanced. Note that the maximum percent, minimum percent and percent range are the same for these two distributions even though the classroom balance pattern is quite different.

Hence maximum and minimum percents and percent ranges are poor indicators of racial balance. The reason they are poor indicators is because they only characterize the extremes of the classroom distribution, and many different distributions, differing greatly in their patterns of racial balance, can have the same extremes.

---

* The AD can also be written

$$AD = \sum_{i=1}^{n} \frac{|p_i - p|}{n},$$

where the Σ indicates the addition shown explicitly in the first equation.

		1	2	3	4	5	6	7	8		P, percent of m-students in school and grade	Minimum percent, m-students in class	Maximum percent, m-students in class	Percent range	Average Deviation percent
1.	m	8	8	8	8	8	8	0	16	32	50	0	100	100	13
	r	8	8	8	8	8	8	16	8	96					
		16	16	16	16	16	16	16	16	128					
2.	m	0	0	0	0	16	16	16	16	64	50	0	100	100	50
	r	16	16	16	16	0	0	0	0	64					
		16	16	16	16	16	16	16	16	128					
3.	m	0	0	0	0	0	0	16	16	32	25	0	100	100	38
	r	16	16	16	16	16	16	0	0	96					
		16	16	16	16	16	16	16	16	128					
4.	m	0	0	0	0	0	0	0	16	16	13	0	100	100	22
	r	16	16	16	16	16	16	16	0	112					
		16	16	16	16	16	16	16	16	128					

Key: m denotes students from minority group m.
r denotes remaining students.

Fig. III-2. Alternative distributions of 128 students to eight classrooms in a given school and grade, and for a particular minority group, m.

The average deviation (AD) looks like a stronger choice for an indicator of racial balance. The AD is much smaller for the first distribution than it is for the second, reflecting the first distribution's greater extent of classroom balance. The difficulty with using the AD as a racial balance indicator is illustrated by the second, third and fourth distributions. All three show the same pattern of classroom balance, specifically complete imbalance, yet the AD decreases in value from 50% for the second distribution to 22% for the fourth distribution. This decline in the AD is a consequence of its dependence on the different racial heterogeneities (shown in the first column of the figure) of the student populations in the second, third and fourth distributions. Therefore the average deviation is a poor indicator of racial balance because it is affected by one of the factors considered to be "uncontrollable" in the previous section of this chapter.

Simple indicators of racial balance suffer from one or both of the faults illustrated by the above examples: either the indicator does not characterize student distributions precisely enough to allow different patterns of racial balance to be distinguished from one another; or part of the variation in the indicator is attributable to one or more of the "uncontrollable" factors identified in the previous section; or both. Therefore the status of racial balance in a

school district can be misrepresented if these simple indicators are used to measure school or classroom balance. If decisions concerning racial desegregation programs are based on the use of simple indicators, such decisions could produce consequences inconsistent with the objectives of racial desegregation.

3. A new indicator of racial balance.

a. A different approach. The previous discussion shows that a different way to characterize student distributions (other than by such simple properties as minimum percents, maximum percents, and so on) must be found if racial balance indicators that meet the requirements of Section A of this chapter are to be developed. The property that suitably characterizes each student distribution, and that leads to acceptable racial balance indicators, is the probability that a particular distribution will occur if students are assigned at random (that is, without knowledge of the racial groups to which the students belong) to schools or to classrooms. The procedure for calculating this probability is well known.* Precisely the same procedure is used to calculate the probabilities (or "odds") of different card combinations in bridge or poker.

---

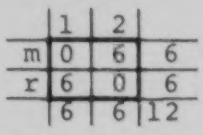
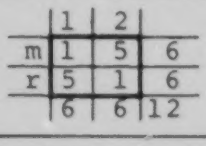
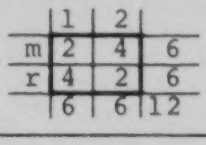
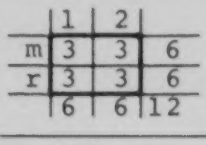
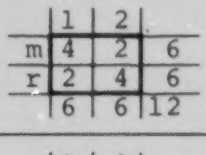
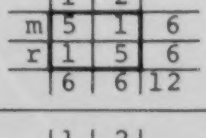
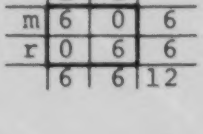
* See any text on the theory of probability, e.g., W. Feller, An Introduction to Probability Theory and Its Applications, vol. I. John Wiley and Sons, New York, 1950.

A simple example will serve to illustrate how a distribution is characterized by its probability of occurrence, and to illustrate how this probability reflects the degree of racial balance inherent in different distributions. Suppose we wish to distribute a student population consisting of 12 students, 6 students from minority group m and 6 students from the "majority" group (labeled r), to two classrooms. Suppose also that each classroom accommodates 6 students (that is, the class size is fixed at 6 students). In this case, there are seven possible distributions. These distributions are shown in Figure III-3, along with the probability that each will occur if the 12 students are distributed, six to each classroom, at random.

The interpretation of these probabilities is quite simple. If the 12 students were distributed to the two classes without any consideration being given to their race, there is only one chance in a thousand (0.1% probability) that distribution 1 and distribution 7 would occur.* In contrast, there are just over four chances in ten (43.3% probability) that distribution 4 would occur. The probabilities of occurrence of the other distributions fall between these two limits, as shown in the figure.

---

* The probability of occurrence of the least probable distribution falls off very rapidly as the student population increases. For example, for 16 m-students, 16 r-students and two classrooms of equal capacity, this probability is approximately 10^{-9}, or 0.0000001%.

Distribution number	Distribution (m denotes minority students r denotes remaining students)	Probability of occurrence	Probability of occurrence expressed as a percent
1.		0.001	0.1%
2.		0.039	3.9%
3.		0.244	24.4%
4.		0.433	43.3%
5.		0.244	24.4%
6.		0.039	3.9%
7.		0.001	0.1%
	Totals	1.001*	100%

* Not equal to 1.000 because of rounding errors.

Fig. III-3. Distribution of 6 m-students (from minority group m) and 6 r-students to two classes, six students to each class.

Distribution 4, the most probable distribution, corresponds to maximum classroom balance: the proportion of minority students in each class equals the proportion of minority students in the population of 12 students being assigned. Distributions 1 and 7, the least probable distributions, correspond to minimum classroom balance: the proportions of m-students in both classes are as far as they can be from the proportion of m-students in the population of 12 students being assigned. Distributions 3 and 5 are less balanced than distribution 4 but more balanced than distributions 2 and 6, and so on.

Thus the probability of occurrence of a particular classroom distribution is directly related not only to the operational definition of classroom balance given in Section A of this chapter, but also to the degree to which the intent of racial desegregation--that assignments to classrooms be made without taking the race of the students into consideration--has been fulfilled. If student distributions of low probability are encountered (i.e., large imbalance), a high likelihood exists that the race of the students was a factor in bringing about those distributions. The lower the probability associated with a distribution, the higher the imbalance, and the higher the likelihood that the assignment

of students was influenced by knowledge of race.* These statements apply not only to distributions of students to classrooms within a school but to distributions of students to schools within a district as well.

The probability of occurrence of a student distribution, unlike the simple indicators discussed earlier, is so intimately related to the extent of racial balance in that distribution, that it is tempting to use the probability itself as a quantitative indicator of racial balance. Unfortunately, the numerical value of the probability is affected by "uncontrollable" factors, such as the size and racial heterogeneity of the student population being assigned, to which a suitable balance indicator must be insensitive.** Therefore,a measure, or function, of the probability must be found from which school and classroom balance indicators satisfying all the requirements discussed in Section A of this chapter can be constructed.

---

* Or influenced by knowledge of some characteristic which is strongly correlated with race. For example, if the black students in a particular district live in a ghetto area, then the place of residence of a student is strongly correlated with the student's race. In this case, assigning students to schools on the basis of where they reside in the district would be equivalent to assigning students to schools on the bases of their race.

** These properties of the probability are discussed in the appendix.

b. Constructing the indicators. In order to formulate suitable indicators, a certain amount of mathematical maneuvering is necessary, beginning with an examination of the formula used to calculate the probability that a particular student distribution will occur. It is not necessary to know specifically how an indicator is obtained in order to use it to calculate school or classroom balance, any more than it is necessary to know how to calculate the odds in poker in order to use the results in playing the game. Consequently, the technical discussion leading to the formulation of the indicators described below has been placed in an appendix to these guidelines, and the discussion in this section has been limited to a description of how the indicators are constructed and interpreted. The formulas required to calculate the indicators and other related quantities appear in footnotes.

The data required to determine numerical values of the school balance indicator for a particular grade (or the equivalent) are the number of students at that grade level from each group (minority and "majority") in each school within the district. The data required to determine numerical values of the classroom balance indicator for a particular grade (or the equivalent) and school are the number of students from each group (minority and "majority") in each classroom (or the equivalent) of the grade. (See Figure III-1 for an example of how classroom data might look.

For each minority group and grade level (or the equivalent), indicators of school balance within the district and classroom balance within each school are separately constructed. However, since the process of distributing students to schools and the process of distributing students to classrooms within schools have a common mathematical representation, the <u>form</u> (although not the meaning) of the school and classroom balance indicators is the same.

• <u>School balance indicator</u>. The school balance indicator for a particular minority group m and a particular grade g will be denoted as B_{mg}. The school balance indicator is expressed as a percent and is constructed as follows:*

$$B_{mg} = 100 \times \left[\frac{(T_{mg})_{max} - T_{mg}}{(T_{mg})_{max} - (T_{mg})_{min}} \right]$$

---

* The quantity T_{mg} which appears in this formula is defined as follows:

$$T_{mg} = \log_2 N - \frac{1}{N}\left[N_m \log_2 N_m + N_r \log_2 N_r + \sum_{s=1}^{S}(N_s \log_2 N_s) - \sum_{s=1}^{S}(N_{ms} \log_2 N_{ms} + N_{rs} \log_2 N_{rs})\right],$$

where:

N = total number of students in grade g in the district (fixed);

N_m = total number of students from minority group m in grade g in the district (fixed);

$N_r = N - N_m$ = total number of students from all other groups (the "remainder") in grade g in the district (fixed);

N_s = total number of students in grade g in school s (fixed);

N_{ms} = number of students in grade g and school s from minority group m (variable);

$N_{rs} = N_s - N_{ms}$ = number of students in grade g in school s from all other groups (variable);

S = the number of schools in the district (fixed).

The quantity $(T_{mg})_{max}$ is the value of T_{mg} which corresponds to the least probable distribution of students from minority group m and grade g to the schools of a district (minimum school balance). For the least probable distribution, T_{mg} takes on its maximum possible value. The quantity $(T_{mg})_{min}$ is the value of T_{mg} which corresponds to the most probable distribution of students from minority group m and grade g to the schools of a district (maximum school balance). For the most probable distribution, T_{mg} takes on its minimum possible value. The quantity forming the denominator of B_{mg}, $\left[(T_{mg})_{max} - (T_{mg})_{min}\right]$, is just the range of possible values which T_{mg} can take on for different distributions of students from minority group m and grade g to the schools of a district.

The quantity $\left[(T_{mg})_{max} - (T_{mg})_{min}\right]$ can be compared to a path to school balance. The path runs from minimum school balance, $(T_{mg})_{max}$, to maximum school balance, $(T_{mg})_{min}$. T_{mg} is a point on this path, and $\left[(T_{mg})_{max} - T_{mg}\right]$ is the distance which the district has traveled along the path to maximum school balance. B_{mg}, then, can be interpreted as the extent of school balance achieved by a district (for students from minority group m and grade g) relative to the extent of balance that the district could achieve, given the size and racial heterogeneity of the student population enrolled in grade g, and assuming fixed school capacities.

• Classroom balance indicator. The classroom balance indicator for a particular minority group m, grade g, and school s will be denoted as b_{mgs}. The classroom balance indicator is expressed as a percent and is constructed as follows:*

$$b_{mgs} = 100 \times \left[\frac{(t_{mgs})_{max} - t_{mgs}}{(t_{mgs})_{max} - (t_{mgs})_{min}} \right]$$

---

* The quantity t_{mgs} which appears in this formula is defined as follows:

$$t_{mgs} = \log_2 N_s \frac{1}{N_s} \left[N_{ms} \log_2 N_{ms} + N_{rs} \log_2 N_{rs} + \sum_{c=1}^{C} (n_c \log_2 n_c) - \sum_{c=1}^{C} (n_{mc} \log_2 n_{mc} + n_{rc} \log_2 n_{rc}) \right]$$

where:

N_s = total number of students in grade g in school s (fixed);

N_{ms} = number of students in grade g and school s from minority group m (fixed);

$N_{rs} = N_s - N_{ms}$ = number of students in grade g in school s from all other groups (fixed);

n_c = total number of students in classroom c of grade g and school s (fixed);

n_{mc} = number of students in classroom c of grade g and school s from minority group m (variable)

$n_{rc} = n_c - n_{mc}$ = number of students in classroom c of grade g and school s from all other groups (variable);

C = the number of classrooms in grade g and school s (fixed).

The indicator of classroom balance, b_{mgs}, can be interpreted in precisely the same way for student assignments to the classrooms of a particular grade and school as the school balance indicator, B_{mg}, was interpreted for student assignments to schools. The indicator b_{mgs} is the extent of classroom balance achieved by school s (for students from minority group m and grade g) relative to the extent of balance that the school could achieve, given the size and racial heterogeneity of the student population to be assigned to the classrooms in grade g in the school, and assuming fixed classroom capacities. The quantities $(b_{mgs})_{max}$ and $(b_{mgs})_{min}$ correspond respectively to the least and most probable distribution of m-students to the classrooms of grade g in school s, and so on.

The construction of maximally balanced and unbalanced school and classroom distributions and the calculations of B_{mg} and b_{mgs} from school and classroom data are intended to be performed on a digital computer. The formulas for T_{mg} and t_{mgs} given above are in a form suitable for machine calculation. Riverside Research Institute staff have performed these calculations for over 20 simulations of student distributions to schools and to classrooms. The results of these calculations, some of which are used as examples later in this section, support the following conclusions: B_{mg} and b_{mgs} are

indicators of school and classroom balance that are consistent with the operational definitions of school and classroom balance introduced earlier; they are independent of one another; and they are not sensitive to factors such as student population size and racial heterogeneity.

Although both B_{mg} and b_{mgs} are quantitative indicators, the calculated values of these indicators must be interpreted with care. The range of values of the indicators B_{mg} and b_{mgs} constitute ordinal measurement scales.* Ordinal scales can only be used to order observed phenomena systematically (in this case, student distributions to schools and classrooms). For example, if two distributions have the same value of B_{mg}, they can be said to represent the same extent of racial balance. Furthermore, a distribution for which $B_{mg} = 50\%$ can be said to show a greater extent of school balance than a distribution for which $B_{mg} = 25\%$. However, it is not correct to conclude that the distribution for which $B_{mg} = 50\%$ reflects twice as much school balance as the distribution for which B = 25%; that is, ratios (and, for that matter, sums and differences) of values of B_{mg} for different distributions have no real significance.

---

* For a discussion of measurement scales, see, for example, S. Siegal, Nonparametric Statistics for the Behavioral Sciences, chapter 3. McGraw-Hill Book Co., New York, 1957.

These limitations should be borne in mind when several school or classroom balance indicators are combined to summarize the status of school or classroom balance throughout a district. Combinations of balance indicators also give rise to ordinal scales, and only relative significance should be attached to the particular values which these combinations may take on.

Particular values of school and classroom indicators may acquire increased significance as a consequence of investigations into the effects of different segregated environments on the cognitive and non-cognitive behavior of children from different minority and "majority" groups. In this event, school and classroom indicators can be adjusted to accommodate the results of such investigations. For example, preliminary data indicate that when only one or two black students are in a classroom, they tend to be more alienated than when the number of black students in the classroom is somewhat larger.* For this reason a school having only a few black students in a particular grade may wish to maintain a degree of classroom imbalance in order to have no less than, say, four black students in any classroom. Under these circumstances, the smallest possible value of t_{mgs} consistent with this constraint would be

---

* Koslin, S., Koslin, B. L. and Pargament, R. Efficacy of School Integration Policies in Reducing Racial Polarization. Paper presented at the meetings of the American Psychological Association, Honolulu, Hawaii, September, 1972.

greater than $(t_{mgs})_{min}$. To reflect this school policy, t_{mgs} corresponding to the distribution in which no less than four black students are found in any class could be defined as "$(t_{mgs})_{min}$." If this adjustment is made, the maximum value of the racial balance indicator would continue to be 100%, corresponding to maximum classroom balance in the face of the policy constraint that no less than four black students must be in a classroom to avoid alienation effects.*

To illustrate the use of the indicators, as well as the discussion in the previous paragraph, the classroom balance in the third grade of a particular school will be calculated. For the purposes of this example, only one minority group will be assumed to be represented (blacks: m=1) and the racial heterogeneity will be assumed to be 10% (i.e., the student population to be assigned to classrooms is 10% black). The remaining conditions specifying the student population to be distributed to the third grade classrooms of schools.

N_s = 240 (total students in the third grade);

N_{1s} = 24 (black students in the third grade);

N_{rs} = 216 (remaining students in the third grade);

C = 8 (number of classrooms in the third grade);

n_c = 30 for all classrooms (class size--the same for all classes in this example).

---

* In presenting values of an indicator calculated from "$(t_{mgs})_{min}$", the constraint should be explicitly identified to avoid misinterpretation.

The distributions required to calculate the classroom balance indicator in this example are shown in Figure III-4. Distribution 1 is the actual distribution of students to the third-grade classrooms of the school. Applying the formula for t_{mgs} to this distribution, $t_{13s} = 0.111$ is obtained. Distribution 2 corresponds to the most probable distribution (maximum classroom balance) for which the formula for t_{mgs} gives $(t_{13s})_{min} = 0$. Distribution 3 corresponds to the least probable distribution (maximum classroom imbalance) for which the formula for t_{mgs} gives $(t_{13s})_{max} = 0.379$. Therefore for the actual distribution (distribution 1), the classroom balance for black students in the third grade in this school is:

$$b_{13s} = 100 \times \left[\frac{(t_{13s})_{max} - t_{13s}}{(t_{13s})_{max} - (t_{13s})_{min}} \right] = 100 \times \left[\frac{0.379 - 0.111}{0.379 - 0} \right] = 71\%$$

On the other hand, if the policy of the school is to have no less than four blacks in each class to avoid alienation effects, then distribution 4 is the appropriate one from which to calculate "$(t_{13s})_{min}$". In this case, the classroom balance for black students in the third grade becomes

$$b_{13s} = 100 \times \left[\frac{(t_{13s})_{max} - t_{13s}}{(t_{13s})_{max} - "(t_{13s})_{min}"} \right] = 100 \times \left[\frac{0.379 - 0.111}{0.379 - 0.044} \right] = 80\%$$

School: s Grade: 3

	1	2	3	4	5	6	7	8	
m=1	7	5	5	7	0	0	0	0	24
r	23	25	25	23	30	30	30	30	216
	30	30	30	30	30	30	30	30	240

$t_{13s} = 0.111$

	1	2	3	4	5	6	7	8	
m=1	3	3	3	3	3	3	3	3	24
r	27	27	27	27	27	27	27	27	216
	30	30	30	30	30	30	30	30	240

$(t_{13s})_{min} = 0$

	1	2	3	4	5	6	7	8	
m=1	24	0	0	0	0	0	0	0	24
r	6	30	30	30	30	30	30	30	216
	30	30	30	30	30	30	30	30	240

$(t_{13s})_{max} = 0.379$

	1	2	3	4	5	6	7	8	
m=1	4	4	4	4	4	4	0	0	24
r	26	26	26	26	26	26	30	30	216
	30	30	30	30	30	30	30	30	240

"$(t_{13s})_{min}$" = 0.044

Key: m=1 denotes black students
r denotes remaining students

Fig. III-4. Examples of distributions required to calculate classroom balance under different conditions.

The value of classroom balance is higher in this "constrained" case because classroom balance is being determined relative to a distribution (distribution 4) in which some imbalance is being maintained in order to avoid alienation effects, rather than relative to the most probable distribution (distribution 2) which corresponds to maximum classroom balance.

When only a small number of minority students are being distributed to schools or classrooms, large changes in the values of the balance indicators can arise if one or two minority students are shifted from one unit to another. For example, suppose the student population contains only one minority student, and suppose further that the entire student population is to be distributed to two classes of unequal size. There are only two possible distributions in this case because the single minority student must be in either one classroom or the other. Necessarily, one of the two possible distributions corresponds to $(t_{mgs})_{max}$ and the other corresponds to $(t_{mgs})_{min}$. Therefore by shifting the single minority student from one classroom to the other, the classroom balance shifts from 0% to 100%, or vice versa.

Situations like this will arise when either the number of minority students is small, or the number of available schools or classrooms is small, or both. In such cases the number of possible distributions, and consequently the number

of different values which the balance indicator can assume, is also small.

Describing racial balance in terms of balance indicators related to the proportions of minority students in schools and classrooms is simply not appropriate when only a few minority students are involved. If the number of possible distributions is, say, ten or less, the balance indicators serve little purpose. The assignment to schools or classrooms of the few minority students involved is best handled on an individual basis in such circumstances, and the use of balance indicators to describe these distributions can be dispensed with.

4. Suggestions for the presentation of school and classroom balance information. Care must be taken to describe the status of racial balance within a district with precision and without distortion. Racial balance is a complex and subtle matter, and poor decisions can result if important elements of school and classroom balance information are suppressed in the decision-making process. The purpose of this section is to suggest ways of presenting racial balance information for decision-making purposes, to suggest methods of summarizing school and classroom information, and to discuss briefly what utility these summaries might have. The discussion is guided by the following precepts:

- School and classroom balance information should be presented separately for each minority group represented in the student population of the district (see the discussion in Section A of this chapter);

- In any presentation, the number of minority students in each school or classroom should be presented along with the corresponding calculated value of the balance indicator. If several grades or schools in a district show a high degree of imbalance, and if district resources for desegregation programs are limited, the number of students affected by different candidate programs could be a determining factor in deciding which imbalance condition to address first. This procedure will also serve to identify grades in which the number of minority students is too small to permit racial balance indicators to be usefully employed.

a. The presentation of school balance information. An example of the suggested format for the presentation of district-wide school balance information for one minority group is shown in Figure III-5. A similar presentation would be made for each minority group attending the schools of the district. In this example, racial heterogeneity for the minority group is 25% for the district as a whole, but it is 45% in the first grade and falls to 5% in the eighth grade. It is assumed that the district has eight schools, and that there are eight

DISTRICT X SCHOOL BALANCE

Minority group: m

Date:

Total minority (m) students: 2,560
Total remaining (r) students: 7,680
Total students: 10,240

Numbers of m-students

grades ↓ — schools →

grades	1	2	3	4	5	6	7	8	m-students per grade	r-students per grade	B_{mg} (%)	$\bar{I}_{mg}$
1	0	20	80	40	96	80	120	140	576	704	68	7.2
2	0	30	60	40	32	100	120	130	512	768	69	6.2
3	0	10	30	30	30	128	130	100	448	832	58	7.4
4	0	0	20	20	30	64	120	130	384	896	53	7.1
5	0	0	0	16	14	40	96	90	256	1,024	56	4.4
6	0	0	0	10	25	25	32	100	192	1,088	61	2.9
7	0	0	0	0	12	16	0	100	128	1,152	36	3.2
8	0	0	0	0	0	0	16	48	64	1,216	28	1.8
m-students per school	0	60	190	156	229	453	634		2,560			
r-students per school	1,280	1,220	1,090	1,124	1,051	827	646	442		7,680		

$\bar{B}_m$ = 59.9%

Key: B_{mg} = school racial balance by grade, m-students
$\bar{I}_{mg}$ = weighted school imbalance by grade, m-students
$\bar{B}_m$ = weighted school balance, all grades, m-students

Fig. III-5. Example of the presentation of school balance information for a particular district.

grades in each school. In practice, a digital computer would generate the presentation.

The array on the left of the figure gives the number of students from minority group m in each grade and school of the district. The usual marginal totals giving the number of m-students per grade (right) and m-students per school (bottom) are shown. The number of r-students (the "remainder") per grade and per school is also shown, set off from the margins of the array with spaces.

For this example, the calculated values of the school balance indicator, B_{mg}, are shown on right side of the figure along with a number called the "weighted school imbalance" ($\bar{I}_{mb}$) which is discussed below. The values of B_{mg} show that school balance is not particularly high in any grade in this district. The largest imbalance exists in grades seven and eight where, however, the numbers of students from minority group m are relatively small.

It is useful to construct a weighted index of school imbalance by grade, $\bar{I}_{mg}$, to flag extreme situations which may require priority action by local administrators.* This index will be large for grades in which school balance is low, or when a large fraction of m-students is affected, and will

---

* $\bar{I}_{mg}$ is defined as follows:

$$\bar{I}_{mg} = \left[\begin{array}{l}\text{weighted school im-}\\ \text{balance by grade}\end{array}\right] = (100 - B_{mg}) \times \left[\begin{array}{l}\text{number of students from}\\ \text{minority group m in}\\ \underline{\text{grade g of the district}}\\ \text{the number of students}\\ \text{from minority group m}\\ \text{in the district}\end{array}\right]$$

be "doubly large" for grades in which school balance is low and the fraction of m-students affected is high.

The values of the weighted school imbalance index for the illustrative case discussed above are shown in the extreme right column of Figure III-5. If district decision-makers are interested in identifying "the largest groups with the greatest needs" as a basis for ordering decision priorities, which is what the imbalance index is designed to do, the imbalance index may be more useful than the school balance indicator. In the example, the third-grade imbalance index is highest, even though school balance is lower in other grades (the seventh and eighth) with fewer minority group students.

If the district wishes to summarize the status of school balance for students from minority group m over all grades, a weighted school balance index, $\bar{B}_m$, can be constructed.*

---

* $\bar{B}_m$ is defined as follows:

$$\bar{B}_m = \left[\begin{array}{l}\text{weighted school} \\ \text{balance, all grades}\end{array}\right] - \sum_{\substack{\text{all} \\ \text{grades}}} \left\{ B_{mg} \times \left[\begin{array}{l}\text{number of students from} \\ \text{minority group m in} \\ \underline{\text{grade g of the district}} \\ \text{number of students from} \\ \text{minority group m in the} \\ \text{district}\end{array}\right] \right\}$$

This index may be useful in making year-to-year or inter-district comparisons. It has the property that its value is heavily influenced by the school balance in grades containing large numbers of minority group students, and less so by the school balance in grades containing fewer minority students. The weighted school balance index for the example shown in Figure III-5 is 59.9%, and appears at the bottom of the Figure.

b. The presentation of classroom balance information. The presentation of classroom balance information requires several separate displays for each minority group attending the schools of the district. A separate display of classroom balance information for each school in the district, analogous to the display of school balance information in Figure III-5, should be presented. A district-wide summary of classroom balance information should also be constructed. An example of a display of classroom balance information for a particular school is shown in Figure III-6. An example of a district-wide summary of classroom balance information is shown in Figure III-7.

Figure III-6 shows classroom balance information for students from minority group m in school number 1 of district Y. Each grade is assumed to have 8 classrooms. Racial heterogeneity is roughly 50% in the school as a whole, but varies from grade to grade. Except for the omission of

CLASSROOM BALANCE

District: Y School: 1 Date:

Minority group: m

grades	Numbers of m-students — classrooms → 1	2	3	4	5	6	7	8	Minority (m) students per grade	Remaining (r) students per grade	b_{mg1} (%)	$\bar{i}_{mg1}$
1	10	12	11	7	10	16	16	16	98	30	71	7.5
2	13	13	3	15	12	16	10	9	91	37	71	6.9
3	15	14	13	15	5	11	11	0	84	44	58	6.9
4	0	14	14	14	5	10	13	0	70	58	46	7.4
5	16	15	16	9	0	0	0	0	56	72	5	10.4
6	12	11	5	8	3	5	5	0	49	79	78	2.1
7	7	7	7	5	5	2	1	1	35	93	87	0.9
8	3	3	3	5	10	2	1	1	28	100	83	0.9
								Totals	511	513		

$\bar{b}_{m1}$ = 60.6%

Key: b_{mg1} = classroom racial balance by grade, m-students

$\bar{i}_{mg1}$ = weighted classroom imbalance by grade, m-students

$\bar{b}_{m1}$ = weighted classroom balance, all grades, m-students

Fig. III-6. Example of the presentation of classroom balance information for a particular school.

marginal totals at the bottom of the array on the left,* the format is identical to the format in Figure III-5.

The calculated values of the classroom balance indicator for school 1, b_{mg1}, are shown on the right of the figure, along with the weighted classroom imbalance index, $\bar{I}_{mg1}$.** The values of b_{mgs} indicate that classroom balance is lowest in the fifth grade, and relatively low in the third and fourth grades. The weighted classroom imbalance index is highest for the fifth grade in this case. Even the larger number of minority students in the third and fourth grades cannot offset the effect of extreme imbalance in the fifth grade.

---

* It makes no sense to combine students in classrooms having the same identifying number in different grades. The classroom number is an arbitrary identifier.

** This quantity is defined in the general case for each grade in a school (by analogy to the school imbalance case) as:

$$\bar{I}_{mgs} = \begin{bmatrix}\text{weighted class-}\\ \text{room imbalance}\\ \text{by grade}\end{bmatrix} = (100 - b_{mgs}) \times \begin{bmatrix}\text{number of students from}\\ \text{minority group m in}\\ \text{grade g of school s}\\ \hline \text{number of students from}\\ \text{minority group m in}\\ \text{school s}\end{bmatrix}$$

The status of classroom balance over all grades in school s for students from minority group m can be summarized by constructing a weighted classroom balance index, $\bar{b}_{ms}$.* This index may be useful in making comparisons among different schools in the district (having the same number of grades) or in making year-to-year comparisons. The weighted classroom balance index for school 1, $\bar{b}_{m1}$ = 60.6%, appears at the bottom of Figure III-6.

Figure III-7 illustrates a district-wide summary of classroom balance information. The district is assumed to have three schools, one of which (school 1) is the school used to illustrate the presentation of classroom balance information in Figure III-6. The array on the left in Figure III-7 gives the number of students from minority group m in each grade and school of the district. The array on the right is composed of the values of the classroom balance indicator for each grade and school within the district. The data show that school 2 has perfect classroom balance in all grades, while in school 3 classroom balance decreases as grade increases. Note that Grade 8 of school 3 is completely unbalanced.

---

* $\bar{b}_{ms}$ is defined as follows:

$$\bar{b}_{ms} = \begin{bmatrix}\text{weighted classroom balance, all grades, school s}\end{bmatrix} = \sum_{\substack{\text{all}\\ \text{grades}}} \left(b_{mgs} \times \begin{bmatrix}\dfrac{\text{number of students from minority group m in grade g of school s}}{\text{number of students from minority group m in school s}}\end{bmatrix} \right)$$

DISTRICT-WIDE SUMMARY OF CLASSROOM BALANCE

District: Y (three schools) Date:

Minority group: m

Total minority (m) students: 1,023
Total remaining (r) students: 2,049
Total students: 3,072

Numbers of m-students

grades	schools 1	2	3	m-students per grade	r-students per grade
1	98	32	32	162	222
2	91	32	32	155	229
3	84	32	32	148	236
4	70	32	32	134	250
5	56	32	32	120	264
6	49	32	32	113	271
7	35	32	32	99	285
8	28	32	32	92	292
m-students per school	511	256	256	1,023	
r-students per school	513	768	768		2,049

Classroom balance

grades	b_{mg1}	b_{mg2}	b_{mg3}	$\bar{i}_{mg}$
1	71	100	100	17.5
2	71	100	96	17.9
3	58	100	90	26.0
4	46	100	82	32.5
5	5	100	70	52.3
6	78	100	46	24.8
7	87	100	30	27.2
8	83	100	0	40.0
$\bar{i}_{ms}$	39.4	0	35.7	

$\bar{b}_m$ = 71.4%

Key: b_{mgs} = classroom racial balance by grade, school s, m-students
$\bar{i}_{mg}$ = weighted classroom imbalance by grade, m-students
$\bar{i}_{ms}$ = weighted classroom imbalance by school, m-students
$\bar{b}_m$ = weighted classroom balance, all grades and schools, m-students

Fig. III-7. Example of a district-wide summary of classroom balance.

The display in Figure III-7 permits rapid association of the degree of classroom balance in any grade and school with the number of m-students affected. For example, the array on the left shows that there are 32 m-students in grade 8 of school 3.

A weighted classroom imbalance index for each grade, $\bar{I}_{mg}$, has been constructed to flag extreme situations.* Local administrators may wish to use these flags as a basis for ordering decision priorities. A weighted classroom imbalance index for each school, $\bar{I}_{ms}$, has also been constructed.** This index could serve to identify schools which might require priority action by local administrators.

---

* $\bar{I}_{mg}$ is defined as follows:

$$\bar{I}_{mg} = \begin{bmatrix}\text{weighted class-}\\ \text{room imbalance}\\ \text{index by grade}\end{bmatrix} = \sum_{\substack{\text{all}\\ \text{schools}}} \left\{ (100 - b_{mgs}) \times \begin{bmatrix}\text{number of students from}\\ \text{minority group m in}\\ \text{grade g of school s}\\ \hline \text{number of students from}\\ \text{minority group m in}\\ \text{grade g}\end{bmatrix} \right\}$$

** $\bar{I}_{ms}$ is defined as follows:

$$\bar{I}_{ms} = \begin{bmatrix}\text{weighted class-}\\ \text{room imbalance}\\ \text{index by school}\end{bmatrix} = \sum_{\substack{\text{all}\\ \text{grades}}} \left\{ (100 - b_{mgs}) \times \begin{bmatrix}\text{number of students from}\\ \text{minority group m in}\\ \text{grade g of school s}\\ \hline \text{number of students from}\\ \text{minority group m in}\\ \text{school s}\end{bmatrix} \right\}$$

These imbalance indices have been calculated for the data shown in Figure III-7. Their values for each grade and school are entered at the right and bottom margins of the array of classroom balance indicators in the Figure. The classroom imbalance index by grade permits the fifth grade to be identified quickly as the grade which probably should receive high priority if desegregation programs by grade are being planned by the district. The classroom imbalance index by school permits school number 1 to be identified quickly as a possible first target if desegregation programs by school are being planned by the district.

For purposes of making the year-to-year or inter-district comparisons described earlier, a weighted classroom balance indicator, $\bar{b}_m$, which summarizes the status of classroom balance for students from minority group m in all grades and schools of the district, can be constructed.* The value of this index for the example given in Figure III-7, $\bar{b}_m = 71.4\%$, is entered at the bottom of the Figure.

---

* $\bar{b}_m$ is defined as follows:

$$\bar{b}_m = \begin{bmatrix}\text{weighted class-}\\ \text{room balance, all}\\ \text{grades and schools}\end{bmatrix} = \sum_{\substack{\text{all}\\ \text{grades}}} \sum_{\substack{\text{all}\\ \text{schools}}} \left(b_{mgs} \times \begin{bmatrix}\text{number of students from}\\ \text{minority group m in}\\ \text{grade g of school s}\\ \hline \text{number of students from}\\ \text{minority group m in the}\\ \text{district}\end{bmatrix} \right)$$

C. The Role of Racial Balance in the Monitoring and Evaluation of Desegregation Programs.

Racial balance indices fill two important roles in the monitoring and evaluation of desegregation programs. First, they are themselves the effectiveness measures which determine the degree to which a desegregation program has been successful in accomplishing its objectives for structural change in the schools and classrooms of a district; or, if a desegregation program extends over several years, they measure the progress which is being made toward the attainment of the desired degree of structural change. Second, racial balance indices, together with racial heterogeneity indices, constitute the minimum set of environmental variables required to determine the cognitive and noncognitive effects of a desegregation program in the second phase of evaluation (see Chapter II).

The data required to determine school and classroom balance indices (as well as racial heterogeneity indices) have been specified earlier in this chapter. It is only necessary to know the number of students from all groups (minority and majority) in each grade in each school, and in each classroom of each grade within each school, to calculate these indices. However, the crux of desegregation lies in the cognitive and noncognitive changes produced in the behavior of participating students, and sufficient

information should be gathered by districts to permit not only the calculation of racial balance indices but also the evaluation of these behavioral changes.

In Chapter V, the major types of evaluative study designs are described in detail. Each requires the collection of different elements of information, but all require more information than is necessary just to calculate racial balance indices. In particular, as indicated in Chapter II, information on student characteristics is required, and baseline data characterizing predesegregation environments is essential.

Not all districts may have the administrative and technical resources to undertake the collection of data required for the evaluation of desegregation programs. Furthermore the calculation of racial balance indices will require computer programming and data processing capabilities which are not available to many districts.

In order to facilitate valid desegregation program evaluations, the State Education Department is prepared to provide assistance to districts in the collection of pre- and post desegregation data required for evaluation studies, and in the calculation, presentation and interpretation of racial balance indices. The State Education Department will generate forms for the collection of the environmental and personal data required for evaluation studies, and will assist

the districts in the implementation of the data collection process.

Using the collected data, the Department will generate the presentations of school and classroom balance information that are described in Section IIIB. These presentations will be provided to the districts, accompanied by analyses which may be helpful in the interpretation of the racial balance information contained in the presentations.

A comprehensive list of the personal and environmental data required to undertake both stages of desegregation program evaluation, and to facilitate interpretations of the results, is given below. Some elements of information are already collected by the Basic Educational Data System (BEDS), and are included in the list for completeness.

1. Information to be collected from each school.

a. School code. School code may be taken from the district's BEDS report.

b. Ethnic or racial composition of professional staff. This is the same as Item #3 on the BEDS school data form.

c. Ethnic and racial composition of the student body. Number (per grade) of white, black, Puerto Rican, and other pupils. This is the same as Item #4 on the BEDS school data form.

d. The proportions of boys and girls (by grade) in the student body.

e. Records of student IQ and prior academic achievement (if measures or indices are comparable for all students).

f. Teacher assignments. Identifies grades where teachers are assigned to the same class all day (self-contained classrooms) as opposed to grades where teachers or children or both change classes during the day.

g. Socioeconomic data. Estimates of the percent of students from low, middle, and high SES (socioeconomic status) homes.

2. Information to be collected from each classroom teacher.

a. School code. The school code would be pre-entered on a form which the teacher fills out.

b. Teacher name. The teacher name can also be pre-entered on the form.

c. Type of assignment. Does the teacher spend the entire day with one class or does she or he have several teaching assignments? In the latter case, does she or he teach a special subject (e.g., art, physical education) where students are assigned to self-contained classrooms with one teacher all day, or does every teacher in that grade or school have multiple assignments? This information

is needed to allow instructional units to be identified and to make sure that such units are not duplicated in balance computations.

d. For each class taught:

1. Grade level

2. Class enrollment list and total number of students.

3. Sex and ethnicity (or race) of each student. Total number of white, black, Puerto Rican, Indian and other pupils.

4. Number of hours per week that the teacher meets with the class.

Chapter IV

Cognitive and Noncognitive Effectiveness Measures for Desegregation Programs

Once it has been determined that implementation of a desegregation program has resulted in the desired change(s) in the racial composition of educational units, the district may proceed to evaluate the effects of the desegregation program on students' cognitive and noncognitive development. Chapter IV is designed to provide districts with guidance in selecting effectiveness measures in both the cognitive and noncognitive domains.

In the cognitive area, where there are many tests and a long tradition of educational measurement, the discussion will focus on two major types of cognitive tests (norm referenced and criterion referenced), with emphasis on some limitations of each type in the evaluation of desegregation programs. In the noncognitive area, where fewer tests exist and where the tradition of measurement (at least in the context of program evaluation) is weaker, the rationale for including noncognitive measures in the evaluation of desegregation programs will be reviewed, criteria for noncognitive effectiveness measures will be outlined, and two racial attitude tests which meet these criteria will be described.

A. Cognitive Effectiveness Measures

In evaluating desegregation programs, cognitive tests are used to determine the effects of the programs on students' achievement in one or more areas of the academic curriculum, e.g., reading, mathematics, science, etc. Achievement in these areas is clearly important to those who manage the schools.

1. Scope of this discussion. An extensive literature exists on the theory and practice of achievement testing. Any attempt to review or even to summarize that literature far exceeds the scope of this chapter.*

Specific standardized achievement tests, such as the Stanford Achievement Tests, the Iowa Tests of Achievement, etc. will not be discussed here either. These tests are already well known to school districts. Moreover, a handbook** designed to assist school districts in selecting standardized tests is now available. This handbook supplements, in a practical way, information provided by test publishers and technical evaluations of tests available in

---

* Interested readers are referred to Cronbach, L.J. Essentials of psychological testing. (3rd ed.) New York: Harper and Row, 1970, and to Thorndike, R.L. (Ed.), Educational measurement, (2nd ed.) Washington: American Council on Education, 1971.

** Hoepfner, R. (Ed.), CSE elementary school test evaluations. Los Angeles: Center for the Study of Evaluation (UCLA), 1970.

reference works.*

This section of Chapter IV focuses on some limitations of two major types of cognitive tests (norm referenced and criterion referenced) as effectiveness measures in the evaluation of desegregation programs, and notes the consequences of these limitations on the district's selection of a design for the evaluation study.

2. Norm referenced tests. The standardized achievement tests familiar to all school districts are norm referenced. A student's test performance is interpreted by comparing it with the performance of other students, usually with the performance of the students on whom the test was originally normed.

A student's raw score on a standardized test (i.e., the number of items answered correctly) has no absolute meaning; it does not directly tell what he knows. A student's raw score acquires meaning only when it is compared with the raw scores of other students. This comparison is usually expressed as a grade equivalent score (grade norm). A grade equivalent score (grade norm) simply identifies the grade level at which a raw score was the mean score of students in the norming sample.

---

* Buros, O. K. (Ed.) The mental measurements yearbooks. Highland Park, N. J.: Gryphon Press, 1941-1965 (irregular).

In evaluating the effects of desegregation on achievement, norm referenced tests assess whether or not desegregation causes a change in the achievement of desegregated students relative to the achievement of other students. The ability to detect such a change will depend both on the test selected and on the group(s) used for comparative purposes.

a. Test selection. All major standardized achievement tests are sufficiently reliable for group comparisons of the type carried out in evaluating desegregation programs; hence, reliability is not an issue in test selection. However, the validity of the test (whether the test measures what the district wants to measure) is a matter of serious concern.

The district must try to select a test whose item content matches its own program reasonably well. If the test does not measure what the schools are teaching, a change in the number of items answered correctly (and hence a change in relative achievement vis-a-vis other students) is unlikely. Even if students are, in fact, learning more after being desegregated, the poorer the match between test content and the district's program, the less sensitive the test is likely to be in detecting achievement changes which result from desegregation.

Item content should also be examined because test names often do not provide an accurate description of what the tests really measure. Striking examples of test items measuring skills other than those indicated by the test title

are found in some reading tests, where items that are supposed to assess reading comprehension in fact measure abstract reasoning skills.* The effects of a desegregation program on achievement in a curriculum area cannot be accurately assessed if items on the test are measuring something else.

In addition to examining item content before a test is selected, a district should examine the test's format and instructions. The effects of a desegregation program on achievement in a subject matter area may be masked if the test format or directions are too complex. Even if desegregation does affect how much students have learned in the area, if students cannot understand test instructions or are confused by test format, they will be unable to demonstrate their higher achievement on the test. In short, an overly complex test format or overly difficult test instructions will probably make the test insufficiently sensitive to detect changes in achievement in the content area. This problem is especially serious in the primary grades.

---

* Since the function of norm referenced tests is to compare the performance of students, these tests are constructed for maximum accuracy in making distinctions between people. Therefore, when a norm referenced test is constructed, final selection of test items is determined largely by statistical properties of the items. The mismatch this may cause between test title and item content is illustrated in Klein, S. P., The uses and limitations of standardized tests in meeting the demand for accountability. UCLA Evaluation Comment, 1971, 2 (no. 4), 1-7.

b. Comparison groups. Assuming that the district succeeds in selecting a test which is valid in terms of test content and format, the ability to detect and accurately assess changes in achievement associated with desegregation will depend on the availability and appropriateness of scores with which to compare desegregated students' scores.

The ability to detect a change in achievement requires, first, that there be base-line data for comparative purposes and, second, that the base-line data belong to a group with which it is reasonable to compare desegregated students' scores.

Without base-line data there are no grounds for concluding that desegregation has or has not had any effect on scores, since there is no way of knowing what students' scores would have been without desegregation. Comparison with national norms (i.e., treating national norms as base-line data) is appropriate only if district students closely resemble the norming sample and typically have test scores similar to those of the norming sample. To the extent that the district's students differ from the norming sample in such characteristics as IQ, SES, curriculum, length of time desegregated, etc., they may be expected to depart from national norms.

Accurate interpretation of changes in district performance under desegregation therefore requires base-line

data from the district itself, either in the form of local norms or control groups. Local norms, developed over a period of years, provide data concerning what scores of students in the district were like prior to the desegregation program, and thus make it possible to detect a change in scores following desegregation. If local control groups are used, base-line data are obtained from groups of students similar to those desegregated but not involved in the desegregation program. With either type of base-line data, comparisons are made between comparable groups, such as between segregated black. ıd desegregated blacks, or between segregated whites and desegregated whites.

3. Criterion referenced tests. In recent years (largely as a consequence of increased need to evaluate the outcomes of individualized instruction) attention has been drawn to a type of test in which a student's performance is directly interpreted with respect to how well he has mastered the subject matter being tested, rather than with respect to how his test score compares with the scores of other students. Tests of this type are called "criterion referenced," since they are built to provide a direct measure of the extent to which students have acquired the criterion skills and knowledge specified by curriculum objectives. Because items on a criterion referenced test are directly related to mastery of specific subject matter, a score on a criterion referenced

test is directly interpretable in terms of what students have learned.

Tests of this type have been routinely used by teachers for many years. A good example is the weekly spelling test used to determine how well students have learned to spell the words studied during a particular week. Students' scores on such tests are interpreted not with respect to how well other students did on the same test in other schools or in previous years, but simply in terms of how much students learned of what the teacher tried to teach during the period of instruction.

In evaluating desegregation programs, criterion referenced tests are used to assess the effects of desegregation on mastery of the curriculum taught in local schools. Comparison groups are required for this purpose. Even though a single criterion referenced score has an interpretable meaning, knowledge of desegregated students' mastery level does not constitute adequate data for evaluation purposes.

In order to demonstrate that desegregation affects test scores, a comparison must be made between the observed (actual) scores of desegregated students and their prior scores or the scores they would be expected to have if they were not desegregated. The district must determine how the number (or proportion) of desegregated students achieving various levels of mastery on the criterion (e.g., 100% correct,

90% correct, 80% correct, etc.) compares with the number (or proportion) of segregated students achieving those levels.*

To derive interpretable data from a criterion referenced test in evaluating a desegregation program, the test must be held constant across the groups of students who are compared.** The requirement for test constancy imposes the further requirement that curriculum content be the same across the groups.

If grade level can be held constant in the evaluation study, it is fairly simple to use criterion referenced tests as effectiveness measures. Under these conditions, if the whole school district undergoes desegregation, scores of students in a particular grade prior to desegregation can be compared with scores of students in that same grade following desegregation (e.g., third graders prior to desegregation are compared with third graders following desegregation). Alternatively, if only part of the district undergoes desegregation,

---

* Comparing numbers (or proportions) of students attaining various levels of mastery is preferable to comparing untransformed mean scores because means are constrained by the "ceiling" of the test, i.e., there is no opportunity for students attaining complete mastery to improve their test scores.

** It is not necessary that exactly the same test be administered each time. Alternate forms of the test, constituted by randomly selecting a fixed number of items per objective, are permissible. They may be used to avoid or solve problems of test security, i.e., to remedy situations in which knowledge of test content leads teachers to "teach for the test."

scores of students involved in the program can be compared with scores of those who are not involved (e.g., third grade program participants versus third grade nonparticipants). In both instances, scores of segregated students are used to provide base-line data for purposes of comparison.

If grade level is not held constant in the evaluation study, criterion referenced tests may still be used as effectiveness measures, but base-line data for all grade levels involved in the study must be collected prior to desegregation. For example, a district may ask: "how does the achievement of third grade students prior to desegregation compare with the achievement of the same students as fourth graders one year later, following desegregation?"

To answer this question, the district must administer the criterion test to students in fourth grade prior to implementation of the desegregation program, so that base-line data are obtained for purposes of later comparison.

Moreover, if the district wants to determine whether desegregation has a cumulative effect on achievement (i.e., whether achievement gets progressively better for students the longer they are desegregated), base-line data must be collected prior to desegregation in all grades where postdesegregation testing will eventually take place.

4. Conclusion. Though norm referenced and criterion referenced tests differ in purpose, construction, and

interpretation, they are alike in that both require the use of local norms or control groups to assess the effects of desegregation on students' academic achievement.

B. Noncognitive Effectiveness Measures

1. Importance of assessing noncognitive program outcomes. In the evaluation of desegregation programs, noncognitive tests are used to assess the effects of the program on students' attitudes. Since the development of more favorable interracial attitudes among students is frequently a goal of desegregation programs, the tests are most often used to determine how interracial attitudes have been affected by the desegregation program. While desegregation may improve interracial attitudes, neither the empirical evidence* nor social psychological theory** suggest that all desegregation programs will necessarily result in such improvement. Therefore, each school district must assess the effects of its own desegregation program on the racial attitudes of its students.

Attitudes should also be assessed because they may play an important role in mediating (i.e., influencing) achievement. In one major educational survey*** it was found that a

---

* Carithers, M.W. School desegregation and racial cleavage, 1954-1970: a review of the literature. The Journal of Social Issues, 1970, 26, 25-48.

** Pettigrew, T.F. Race and equal educational opportunity. Harvard Educational Review, 1968, 38, 66-76.

*** Coleman, J.S., et al. Equality of educational opportunity. Washington: U.S. Government Printing Office, 1966.

combination of several attitudinal factors was better able to predict students' verbal achievement scores than were all school factors (e.g., school facilities, teacher characteristics, per pupil expenditures, etc.) combined. The survey also suggested that interracial attitudes may be a key variable in determining the effects of desegregated school environments on black achievement. It was found that those black students reporting the greatest proportion of close white friends also had the highest achievement scores. Furthermore, black achievement was best in schools reporting the smallest amount of interracial conflict. Thus, in evaluating desegregation programs, assessment of racial attitude changes may facilitate understanding of changes in academic achievement.

Districts may also wish to use noncognitive measures to assess program effects on other attitudes, such as self-concept, achievement motivation, and sense of fate control, which are important in personal-social development and which may affect achievement.

2. Criteria for noncognitive effectiveness measures. To function as an effectiveness measure in the public schools, a test should be group administrable in the elementary grades, should require low levels of reading skill and verbal comprehension, should be relatively resistant to socially compliant

responses, should be objectively scorable, should be reliable, and should have demonstrated validity for purposes of program evaluation.

Very few noncognitive tests have been developed which meet all the above criteria. The discussion in this section of Chapter IV focuses on the People Test and the Pick-A-Class Test, two currently available measures of interracial attitudes which do meet the criteria. These tests may be obtained through the State Education Department, which will (at the district's request) aid in scoring the tests, carry out computer analyses of the data, and interpret test results for school districts.

3. The People Test: A measure of social distance. The People Test is a nonverbal measure of students' social distance attitudes and beliefs. Social distance refers to the degree of intimacy or closeness in interpersonal interactions which members of one group are willing to share with members of another group. Social distance stems from the distinctions which a culture makes between people with respect to such dimensions (characteristics) as race, sex, age, ethnicity, etc., and is believed to be a universal phenomenon.

The People Test is specifically concerned with the social distances which students attribute to the socially important dimensions of race and sex. The principal use

of the People Test in evaluating a desegregation program is to determine whether and to what extent the program results in a change in these distances. Because patterns of racial prejudice in the United States traditionally have had important sexual components, the People Test also assesses sexual social distances and racial-sexual social distances.

a. Method and format. Social distance concepts are assessed by asking that students judge the relative degree of similarity between figures (drawings) of children who differ by race, by sex, or by both race and sex. Students are instructed to express the degree of similarity between any two figures by means of the amount of space they place between those figures. It is assumed that in making these judgments the child transforms the distinctions he has learned to draw between people into metric distances, and that the smaller the distances placed between any two figures, the smaller the social distance between the people represented by those figures. The distances employed in the test are not assumed to have any absolute meaning in terms of feet, yards, etc.; rather, it is assumed that their relative size reflects the comparative importance of race, sex, and race-sex as characteristics by which distinctions are made between people.

The figures used in the test are line drawings of a white boy, a white girl, a black boy, a black girl, and a

stick figure representing the self. These figures are age-appropriate for students taking the test (there are three versions of the test, one each for grades 1-4, 5-8, 9-12).

Students are instructed that the greater the degree of similarity between any two figures, the closer together the figures should be placed; the smaller the degree of similarity between the figures, the further apart they should be placed. One figure is printed on a pressure sensitive tab, enabling students to paste it at whatever distance they wish from the other figure, along the response scale provided.

b. Two types of test items. In some test items, students judge the similarity between two figures (excluding the self figure) which differ from each other by race, by sex, or by race and sex. Responses to these test items reflect students' concepts concerning the racial and sexual distinctions which are ordinarily made between people in our culture. These items thus assess beliefs concerning the socially standardized or normative distances between the races and the sexes.

In other test items, students judge the similarity between themselves (self figure) and figures of children who differ from themselves by race, by sex, or by race and sex. Responses to these test items reflect students'

concepts of the relative importance of race and sex differences between themselves and each of the other figures. These items thus assess the degree to which personal attitudes concerning distance from others vary as a function of the race and/or sex of the other person.

c. Administration. The People Test is group administered and requires approximately 15 minutes for students to complete. Students need not be able to read at all in order to take the test.

The test may be administered by classroom teachers or by other school personnel. Test administration involves no special skills, other than a careful reading of the administrator's manual prior to the testing session. About 15 minutes are required for the administrator to familiarize himself (herself) with the manual.

During the testing session the administrator uses the manual to guide students through the test. He (she) identifies the figures, explains and illustrates how distance along the response scale is used to represent the degree of similarity between figures, and helps students to complete one or two practice items. Students then complete the test on their own.

d. Scoring. Scoring of the People Test is completely objective, and consists of recording the amount of distance placed between the two figures in each test item. Since a

metric "ruler" is printed beneath the test items, the task of recording the distances between figures is a simple one: the scorer need only look below the figure on the pasted tab to find the score for that item. Scores for each item may either be recorded on a roster sheet for later keypunching or may be keypunched directly from the test booklet.

e. Analysis of test data. People Test data may be analyzed to determine whether (and to what extent) the desegregation program affects the distances which students place between the races and sexes (social distance beliefs) as well as to assess the program's effects on personal distances from peers who differ by race, by sex, or by race and sex (social distance attitudes). Data may also be analyzed to determine whether desegregation alters the extent to which a student's social distance beliefs and attitudes correspond.

Like scores on norm referenced cognitive tests (discussed earlier in this chapter), People Test scores have no absolute meaning; they are not directly interpretable in terms of what students think or feel. Interpretation of students' People Test scores requires a comparison with the scores of other students. Therefore, local norms (test scores prior to desegregation) or control group data (scores of students who have not been desegregated) are needed to determine program effects on People Test scores.

A total score is not calculated for the People Test. Instead, analyses are based on students' responses to individual test items.* Before substantive analyses are

---

* One way of structuring these analyses is by examining the responses of all students to any given test item. For those test items which do not include the self figure (i.e., for the normative distance judgments, such as the distance between a white girl figure and a black girl figure, or between a white boy figure and a black boy figure, etc.) this type of analysis yields interpretable data, since the item has a reasonably constant meaning for all students.

However, for those test items involving the self figure (i.e., for personal distance judgments, such as distance of self from black boy, distance of self from white girl, etc.), the meaning of any item varies according to the race and sex of the student who is taking the test; thus interpretation of responses to any given test item is difficult. Accordingly, a procedure has been developed for analyzing personal distance judgments so that the results will have a common meaning for all students.

In this procedure, the test items used in analyses vary according to the race and sex of the student taking the test and according to the type of personal distance being studied. For example, to study personal distance from opposite race peers of the same sex, the pertinent data for students who are black boys would be distance scores on the self-white boy item; the pertinent data for students who are black girls would be distance scores on the self-white girl item, and so forth. Following the same logic, it is also possible to analyze other personal distances (e.g., distances from peers of opposite sex but same race, and distances from peers of opposite race and opposite sex). Analysis of variance is the statistical procedure most often used in carrying out these analyses.

It is also possible (by covariance analyses) to statistically remove from personal distance judgments the influence of normative distance judgments. In addition to providing a purer measure of attitude than is obtained from a simple analysis of personal distances, covariance procedures also effectively control possible student tendencies to restrict judgments to particular segments of the response scale.

begun, a computer program is used to identify and delete any students whose pattern of responses suggests that they have not followed test directions.*

f. Validity. Because the People Test is presented as a cognitive task (i.e., students are asked to judge how alike or similar two people are), the actual purpose of the test as a measure of attitude is somewhat disguised. To the extent that the disguise is successful, the test will not induce student anxiety about revealing attitudes and the likelihood that students will give socially desirable (rather than "true") responses will be reduced.**

---

* While students are free to use any distance they choose for each test item, it is assumed that if they understand and follow directions the magnitude of their judgments will vary across the items. Therefore, all test data of students showing little or no variability in responses across items are deleted from analyses.

** However, even if some students do recognize the test as an attitude measure, it is more difficult to deliberately distort responses on the People Test than on most other attitude scales. Students who try to disguise their attitudes by the simple device of pasting the tabs at the same distance for all test items are easily detected and deleted when the data are initially screened. If students attempt to distort their responses while varying their distance judgments from item to item, the fact that the figures vary simultaneously in two dimensions makes it very hard to intentionally distort responses in one dimension (e.g., race) without simultaneously distorting responses in the other dimension (e.g., sex). The fact that both the race and sex dimensions are consistently recovered when People Test data are analyzed (see below) indicates that most students do respond honestly to the task. While it is theoretically possible for students to alter responses in one dimension without impairing the integrity of judgments for the other dimension (thereby "faking" the test without being detected), few students could be expected to formulate and follow the complex rules which would be required to do so.

Several studies have provided evidence concerning the construct validity of the People Test. It has consistently been found* that students systematically utilize the race and sex dimensions in the stimulus figures when making their judgments of personal and normative social distance.**

One important construct validity study*** investigated developmental changes in People Test normative distance scores for a sample of over 4,000 black and white students in grades 1-12. The developmental data were analyzed to determine whether they were in accord with known socialization patterns in the United States. It was found that

---

* See, for example, Koslin, S., Koslin, B. L., Pargament, R. and Waxman, H. Classroom racial balance and students' interracial attitudes. Sociology of Education, 1972, in press.

** Among the normative distance judgments, those test items which involve figures differing in two dimensions always receive larger distances than those test items where the figures differ only in one dimension. In judgments involving the self, distances are clearly related to the degree of similarity between the characteristics of the test-taker and the characteristics of the target figure. Boys consistently place the self figure closer to male stimuli, whereas girls place the self figure closer to female stimuli. Black students place the self closer to black than to white target figures, whereas white students do the opposite. When multidimensional scaling procedures have been used to analyze the data, it has been found that students appropriately locate the self figure in the "space" created by the shared properties of the other stimuli.

*** Koslin, S., Koslin, B. L., Pargament, R. and Bird, H. Children's social distance constructs: a developmental study. Paper presented at the meetings of the American Psychological Association, Washington, D.C., September 1971.

social distance between the races increases with age and is greater between opposite race girls than between opposite race boys. Moreover, it was found that social distance between the sexes decreases following a preadolescent rise, and is larger between whites than between blacks. In general, it was found that black students attribute greater social distance to racial differences than white students do. All these findings (as well as others too complex to review here) correspond sufficiently well to known socialization patterns in our culture to provide strong support for the construct validity of the People Test.

Another validity study showed that People Test scores are related to scores on other racial attitude tests.* Personal distances were compared for students who had shown extreme pro-white or pro-black preferences for teachers and classmates on a test of racial preferences. It was found that students with pro-white preferences placed significantly less distance between the self and the white girl figure than did students with pro-black preferences; students with strong pro-black preferences placed significantly less distance between the self and the black girl figure than did students with pro-white preferences. Corresponding trends appeared for male stimulus figures. Thus it was concluded that

---

* Koslin, S. The measurement of schoolchildren's racial attitudes: a validity study. Paper presented at the meetings of the Eastern Psychological Association, Philadelphia, 1969.

students' personal distances from peers of the opposite race are significantly related to their preferences concerning classroom racial composition.

The People Test has been used in several studies designed to investigate the effects of student assignment policies on interracial attitudes. In one study* it was found that normative and personal distances are smaller in schools with balanced than in those with unbalanced classes, especially when the figures being judged differ by both race and sex. In a second study** it was found that personal racial distances are smaller in racially mixed (heterogeneous) schools than in racially isolated (homogeneous) schools within the same district. A third study*** found that when the proportion of blacks in a class drops below a certain level, the personal distance of black students from opposite race peers increases.

---

* Koslin, S., Koslin, B. L., Pargament, R. and Waxman, H. Classroom racial balance and students' interracial attitudes. Sociology of Education, 1972, in press.

** Koslin, S., Amarel, M. & Ames, N. A distance measure of racial attitudes in primary grade children: an exploratory study. Psychology in the Schools, 1969, 6, 382-385.

*** Koslin, S., Koslin, B. L., and Pargament, R. Relationships between educational integration policies and students' racial attitudes. Paper presented at the meetings of the American Educational Research Association, New York City, February, 1971.

Thus, in policy-related studies to date, the People Test has proved sensitive to differences in classroom racial balance, school racial heterogeneity, and classroom racial heterogeneity.

g. Reliability. Test-retest studies have shown that the People Test is adequately reliable for analyses of group data. Reliability data are based on test-retest studies in which alternate forms of the People Test were administered to approximately 600 students in grades 3, 5, and 9, with a one-month time interval between testings. While it was found that item reliabilities vary somewhat according to the type of item and age of subjects,* they are high enough for group comparisons of the type carried out in the evaluation of desegregation programs.

4. The Pick-A-Class Test. The Pick-A-Class Test is a semi-disguised measure which assesses children's racial preferences for classmates and teachers. In evaluating the effectiveness of a desegregation program, the Pick-A-Class Test is used to measure differences between the classmate

---

* Items measuring personal social distances are more reliable than those measuring normative social distances (average item reliability is $\bar{r} = .63$ for personal judgments and $\bar{r} = .42$ for normative judgments). Item reliabilities are higher for fifth and ninth graders than for third graders (average item reliabilities in those grades are, respectively, $\bar{r} = .56$, $\bar{r} = .56$, and $\bar{r} = .39$). Since these reliability coefficients are based on individual test items, rather than on total test score, they can be considered acceptably high for group comparison purposes.

and teacher preferences of desegregated students and those of segregated students, or to assess differences between students' preferences prior to and following desegregation. The test is designed for use in grades 1-4.

a. Format and method. The test materials are 18 sketches of classroom scenes. The sketches vary systematically in the type of activity portrayed (three activities are shown), in the race of teacher portrayed (two variations: black and white), and in the racial composition of pupils portrayed (three variations: all white, all black, racially mixed).

The 18 sketches are organized into nine pairs so that all possible comparisons are made between the different racial compositions, with classroom activity counterbalanced. On any page, two of the six racial compositions appear, each embedded in a different classroom activity. In some comparisons teacher race is held constant while pupil race and classroom activity vary. In other comparisons pupil race is held constant while teacher race and classroom activity vary.

On each page of the test booklet, students indicate with a check mark the class that they would prefer to be in. After they have indicated their preference on all pages, students go through the entire test booklet again and on each page indicate by means of an "X" the class which they would prefer not to be in.

b. Administration. The Pick-A-Class Test is group administered and requires approximately 10 minutes for students to complete. Students need not be able to read at all to take the test.

The test may be administered by classroom teachers or by other school personnel. Test administration requires no special skills, other than a careful reading of the brief administrator's manual prior to the testing session. About ten minutes are required for the administrator to familiarize himself (herself) with the manual.

During the testing session the administrator uses the manual to illustrate at the chalkboard what students need to do, and to guide them through the first two pages of the test. Students then complete the rest of the pages in the test booklet on their own. The entire procedure is repeated when students go through the test the second time (to mark on each page the class that they would prefer not to be in).

c. Scoring. The test is scored by first recording which sketch was selected as preferred and as not preferred on each page. This information may be rostered for later keypunching or may be keypunched directly from the test booklet. Students who have selected the same sketch as preferred and as not preferred are detected and deleted at this point.

On the basis of his pattern of responses, each student is assigned a score for teacher preference, a score for peer preference when a white teacher is portrayed, and a

score for peer preference when a black teacher is portrayed.*

d. Data analyses. Like scores on criterion referenced tests (discussed earlier in this chapter), scores on the Pick-A-Class Test are directly interpretable. However, to determine whether desegregation affects test scores, comparisons must be made between the scores of desegregated students and their prior scores or the scores they would be expected to have if they were not desegregated. Local norms or control groups are required for such comparisons.

Separate analyses are carried out for students' preferences concerning teacher race, classmate race when a white teacher is portrayed, and classmate race when a black teacher is portrayed.**

e. Validity. Since the various racial compositions of the Pick-A-Class Test are embedded within different classroom activities on each page, it is difficult for the student to infer which aspects of the stimulus sketches are considered important by adults. This partial disguise of the purpose of the test should reduce the likelihood that students will give

---

* For teacher preference, the student is scored as being either pro-white, pro-black, or as having no consistent racial preference. For classmate preferences, the student is scored as being either pro-white, pro-black, pro-mixed, or as having no consistent racial preference.

** Since scores on the Pick-A-Class Test constitute nominal rather than ordinal data, nonparametric statistics such as chi-square are used in analyses.

socially compliant responses. Even if some students do infer the purpose of the test, they are never confronted with having to choose between two pictures varying only in racial composition: for those who prefer not to choose on racial grounds, there is the opportunity to respond in terms of the activities portrayed.

The construct and predictive validity of the Pick-A-Class Test have been demonstrated in several studies. It has been found that students in schools with racially balanced classes show significantly less racial polarization in their preferences for teachers, and somewhat less racial polarization in their preferences for classmates, than students in schools with racially unbalanced classes.* In a study of the effects of classroom racial heterogeneity on attitudes (Koslin, Koslin, and Pargament, 1971),** it was found that black students in classes with 15% or fewer blacks are more likely to express a preference for all black classmates than are black students in classes with about 50% blacks. Finally, in a study carried out to determine whether

---

* Koslin, S., Koslin, B.L., Pargament, R., & Waxman, H. Classroom racial balance and students' interracial attitudes. Sociology of Education, 1972, in press.

** Koslin, S., Koslin, B.L., and Pargament, R. Relationships between educational integration policies and students' racial attitudes. Presented at the meetings of the American Educational Research Association, New York City, February, 1971.

scores on an attitude test battery which included the Pick-A-Class Test could predict sociometric choices, classmate and teacher preference scores showed higher correlations (r = .28 and r = .22) with the proportion of black peers nominated than did any other predictor variable.

Thus studies have shown that the Pick-A-Class Test is sensitive to differences in classroom racial balance and classroom racial heterogeneity, and that it predicts sociometric choices.

f. Reliability. The Pick-A-Class Test was administered to approximately 200 third grade students in a test-retest reliability study with a one-month interval between testings. The majority of students received exactly the same score on the posttest as they had on the pretest,* and only two percent showed a reversal in classmate or teacher preferences. These data indicate that the Pick-A-Class Test is sufficiently reliable to detect differences between groups,

---

* For teacher preference, 70% of students had identical scores on the pretest and on the posttest (χ^2 = 71.3, df = 4, p < .001). For classmate preferences in comparisons with black teachers, 58% of students received the identical score in the pretest as in the posttest (χ^2 = 93.9, df = 9, p < .001); for classmate preferences in comparisons with white teachers, 55% of students received identical scores on the pre- and posttests (χ^2 = 87.6, df = 9, p < .001). Based on these data, reliability boundaries for the teacher preference subtest are estimated as .57 < ρ < .75; for classmate preferences with a black teacher, .45 < ρ < .68; and for classmate preferences with a white teacher, .41 < ρ < .64.

and thus is suitable for the types of between group analyses carried out in the evaluation of desegregation programs.

CHAPTER V

Evaluation Designs

This chapter reviews some of the major design issues which school districts are likely to encounter in carrying out studies to determine the cognitive and noncognitive effects of desegregation programs. Some of the design issues discussed in this chapter are peculiar to the evaluation of desegregation programs and are reviewed in detail because they may not be familiar to school district personnel. Other problems discussed are more classical issues of evaluation design which may be familiar to the reader but which, because of the special importance they assume in the context of desegregation research, warrant mention here.

Solutions to some of the problems discussed are likely to be beyond the capability and resources of local school authorities, and will require that the district seek professional assistance from the State Education Department (SED) or from qualified consultants. It is the purpose of this chapter to signal these problems for school district personnel so that they will be able to distinguish between problems which they can handle on their own and those with which they will require assistance. Moreover, since the nature of the problems, e.g., statistical, administrative, etc. will be

specified, districts should be able to define the particular types of consultative skills and expertise which they will need.

This chapter should also alert districts to the possibility that there may be one or more problems of evaluation design which, given the nature of the district and the program which is planned, cannot readily be resolved completely even with outside help, and which will therefore require that effectiveness test data be interpreted cautiously.

Organization of this chapter. The heart of this chapter is the presentation of alternate types of evaluation designs and a discussion of their strengths and weaknesses. Before these are presented, a number of theoretical and procedural issues which cut across all designs are discussed. Once the concepts embodied in these issues are understood, the specific strengths and weaknesses of the particular designs should be more easily grasped.

In view of the multiplicity of existing and potential desegregation programs, no attempt will be made to present a comprehensive review of all possible evaluation designs. Instead, a description of some principal design types and an analysis of their associated costs and benefits should suffice to sensitize readers to the major design issues.

The various types of designs will be presented sequentially, from the simpler to the more complex. However, in

grouping the designs for a logically ordered discussion, it will not be useful to employ a single simple-to-complex progression, since desegregation programs which involve only a portion of a district's students require a fundamentally different design approach than programs where all students are involved. Therefore, designs for the two types of situations (all students involved versus only some students involved) will be considered separately. While the actual design details (and thus the information derived) differ between these two major classes of programs, it will be seen that the two types share a largely overlapping set of basic methodological concerns.

It will further be seen that in desegregation research the simpler designs usually have limited utility with respect to yielding definitive conclusions about the effects of a desegregation program and that, as increased design and analytic complexity are added to the simpler plans, the information gained and clarity of interpretation increase. However, each gain derived from increased complexity usually comes at some type of cost to the district. These costs and benefits will be summarized at the end of the chapter.

A. Issues Which Cut Across All Designs

1. Selection of independent and dependent variables. All designs to be discussed share a common purpose: to assess the effects of varying selected environmental factors on one or more kinds of cognitive and noncognitive outcomes, given the context of a particular set of student and school characteristics. The care with which these various factors are represented in the research design will, in large measure, determine the interpretability and utility of the evaluation study.

a. Environmental variables. The same environmental factors (e.g., racial balance, racial heterogeneity, etc.) which were treated as dependent data (outcomes) in the analysis of structural changes during the first stage of evaluation are used as independent or "control" variables (i.e., as classificatory measures) in assessing the cognitive and noncognitive effects of desegregation.

The environmental characteristic directly altered by the desegregation process (classroom racial balance, school racial heterogeneity, or any other structural characteristic) constitutes the principal independent variable. Using this variable to classify students enables the district to ask the question: What consequence does a change in the level or value of this variable (e.g., balanced as opposed to unbalanced classes, high proportion of black students versus low proportion of black students in different schools, etc.) have on cognitive or noncognitive outcomes?

To obtain an accurate answer concerning the effect of a desegregation program on cognitive and noncognitive outcomes, those balance and heterogeneity characteristics which are not affected by the desegregation program but which nevertheless vary across educational environments must be accounted for in the design. Variation in these factors has either been shown by research (or may be presumed on strong theoretical grounds) to affect cognitive or noncognitive outcomes. Consequently, the desegregation treatment may not have the same effects under different environmental conditions, i.e., the effects of desegregation may interact or vary with the surrounding conditions.

Since the design should permit the district to determine whether the effects of desegregation vary as a function of other environmental conditions, it will be necessary to hold constant or otherwise control these other conditions. The object of such control procedures is to remove the effects of these other factors so that, as nearly as possible, the effects of the structural change produced by the desegregation program can be assessed.

When the balance and heterogeneity indices show that a particular factor has a constant (homogeneous) value in all environments, then that factor is, in effect, naturally "controlled" and there is no reason to use it as an independent

or classificatory variable. For example, if in a district where the desegregation objective is to change school racial heterogeneity all schools are found to have racially balanced classes both before and after desegregation, there is no reason to use classroom racial balance as a variable in the design. Since classroom racial balance does not vary, it cannot have any measurable effect, i.e., it cannot differentially affect students' scores.

However, if a balance or heterogeneity factor other than the one deliberately being altered by the program is found to vary across environments, it should be used as a way to classify students for the evaluation design. For example, in a desegregation program designed to affect school racial balance, some students may be assigned to schools with racially balanced classrooms, whereas others may be assigned to schools with racially unbalanced classrooms. (This could occur if some schools in a district practiced ability or achievement grouping whereas others did not.) Since classroom racial balance can affect students' attitudes and achievement, districts in which different schools vary widely in degree of classroom racial balance should classify students according to the level of classroom balance experienced, so that comparisons can be made to determine the effects of different levels of school racial balance under racially balanced versus racially unbalanced classroom conditions.

To take another example: if the objective of the desegregation program is to alter classroom racial balance within schools, and if there happens to be appreciable variation between schools with respect to racial heterogeneity (i.e., the schools are unbalanced), this heterogeneity factor should be used as a variable in the design, so that it will be possible to detect and compare differences in the cognitive or noncognitive effects of altering classroom racial balance when the proportion of black students is low versus when it is high.

Another important consideration in the selection of independent and control factors is the possibility that a change in one environmental variable may be accompanied by a change in a second variable. When two or more variables are correlated with each other, and hence change simultaneously, it is difficult to interpret either variable as a fully independent factor in the design.

For example, because of the correlation in our society between race and social class (SES), a planned change in racial balance is often accompanied by a change (planned or unplanned) in SES balance. Since it is widely recognized that SES influences achievement, the correlation between the race and SES composition of educational units makes it difficult to assess the specific cognitive or noncognitive effects attributable to a change in racial composition.

Because of the strong race-SES correlation in the general population, most school districts cannot reasonably expect to disentangle the effects of changes in SES milieu from those of changes in racial composition. However, in some districts it may be possible to design studies so that racial composition and SES composition are distinguishable as factors, in order that the relative influence of changes in racial versus SES composition can be evaluated. This could happen if some black students were randomly assigned to high SES schools and others were randomly assigned to low SES schools. Comparisons between blacks assigned to classes or schools differing in SES composition could provide information on the effects of SES desegregation as compared with the effects of racial desegregation (provided that the groups of blacks assigned to the two schools were initially equivalent and that curriculum, teachers, special services, etc. were comparable in the high and low SES schools).

As was pointed out in Chapter II, various processes may be used to implement desegregation. The effects of desegregation may vary according to the process used. In selecting environmental factors to serve as independent or control variables in an evaluation design, districts should treat implementation processes as a type of environmental variable and should select a design in which various implementation

processes can be distinguished whenever several processes are being used simultaneously.

For example, in implementing desegregation, the district may assign some black students to schools within walking distance of their homes whereas others may be bused to another part of the district. Under such circumstances, program effects may differ between bused and non-bused students since, for example, opportunities for casual and informal interaction between the races will probably be fewer for students who are bused than for those who are not; hence interracial attitudes may differ for the two groups of black students. In this example, mode of transportation should be considered as an independent or control variable, so as to permit comparisons to determine whether program effects are different for walkers versus riders.

School districts may also find that certain environmental variables are correlated, e.g., that variation in classroom racial balance tends to be associated with variation in proportion of black students, or that proportion of black students tends to be associated with busing versus walking, etc. Under such conditions the variables are not fully independent of each other (i.e., they are "confounded"), and unambiguous data will be hard to obtain. A consultant with considerable skill in experimental design and statistical analysis will be required to suggest a design which will make it possible to

separate, and thus to assess, the effects of these empirically confounded (correlated) factors,at a cost that the district can afford.

b. Personal variables. A number of personal student characteristics should also be used as variables in the design, since the effects of the desegregation program may differ for various subsets of the student body. For example, the program may not affect blacks and whites similarly, or may have a different effect on high versus low SES students, etc. The evaluation design should enable the district to assess any such differential program impact. The most frequently used personal variables for classifying students are: race, age (grade), sex, IQ, achievement level, and SES background. The more detailed the conclusions which a district wishes to draw about program effects on different groups of students, the greater the number of personal variables which must be represented in the design.

2. Independent variables must be limited in number. The discussion to this point has illustrated the need for careful review of all the environmental and personal variables which might affect the outcomes of desegregation, and has pointed out that a district's ability to draw precise conclusions will depend on the selection of appropriate independent variables for the evaluation study.

However, no district is likely to be able to treat all relevant factors as independent variables. First, an evaluation study rapidly becomes both administratively unmanageable as well as prohibitively expensive as independent variables are added, because the sample size required for reliable data increases geometrically as the number of independent variables or the number of levels of the independent variables is increased. For example, increasing the number of grade levels tested from one to two doubles the size of the required sample; treating both classroom racial balance and classroom racial heterogeneity as independent variables requires twice the sample size needed when only one of these variables is used.

Second, the confounding of certain variables (as illustrated above) creates statistical problems in designs, since not all possible combinations of the independent variables and their levels will actually be found in the district. For example, racially balanced classrooms may be found only at some levels of racial heterogeneity, rather than at all levels. Under these circumstances, only one or the other of the balance and heterogeneity factors may be used as an independent variable.

Finally, the sheer analytic complexity of data from designs with more than four independent variables probably exceeds the available interpretive capability of most districts and of most consultants. Working with a statistical consultant, district personnel should therefore review the environmental

and personal variables which are particularly important in their district, in order to select the major program treatments which must be used as independent variables and to identify other important factors which can be treated as control variables via any of several sampling or statistical procedures (e.g., nested designs, replicate designs, etc.).

3. Dependent variables. In the assessment of cognitive and noncognitive effects of desegregation, the dependent variables are scores on those tests selected by the district to serve as effectiveness measures, e.g., attitude test scores, reading test scores, arithmetic test scores, etc.

Each additional dependent variable included in the evaluation study increases the variety of cognitive or noncognitive outcomes assessed, but also adds to the costs of the study with respect to test purchase, length of testing session, scoring of booklets or answer sheets, computer time, and interpretation of results. However, in contrast with the costs of increasing the number of independent variables or their levels, the number of dependent variables does not affect sample size or characteristics.

4. Advance selection of variables and planning of designs. It was noted in Chapter II that if analysis shows that the desired structural changes in student body composition have not taken place, the district should not proceed to assess cognitive or noncognitive program effects, since the planned desegregation has not actually occurred.

Despite this need to verify the structural consequences of desegregation programs before proceeding to assess cognitive and noncognitive outcomes, the design for the latter study should be decided upon before desegregation is begun. The design chosen will determine whether or not students need to be tested before desegregation is implemented* and, if so, who will be tested. Specification of the design may also influence assignment policies, so that students can be retained in classifications needed to complete the study. For example, if level of racial heterogeneity is a control variable, students should be assigned so that they are in the same level after desegregation as before.

Therefore, it will usually be preferable to fully design the evaluation study ahead of time, rather than to defer design decisions until the extent of the structural change brought about by desegregation has been determined. If necessary, elements of the design or analytic procedures can be modified on the basis of unanticipated changes shown by the balance and heterogeneity indicators.

5. Sample size. The number of students to be tested, as well as their characteristics, will vary as a function of the independent variables selected for inclusion in the design,

---

* The discussion below will show that most types of designs do require testing prior to desegregation.

the size of the district, and the financial resources available for the study. However, for any design there will be a minimum sample size required. Below this minimum sample size, there would be too few students representing the levels and combinations of the independent variables to yield stable data. Therefore, the school district should work with its statistical consultant to determine what the minimum number of students of various characteristics must be to permit the desired analyses to be performed.

Where resources are limited, and there is consequently a ceiling on the total number of children who may participate in the study, the district must be especially careful not to spread the sample across so many grade levels that there will be an inadequate number of students in each. While it may be desirable to test at more than one grade level so that the effects of desegregation at different ages can be assessed, it will not be worthwhile to do so if the resulting number of students in each grade will be too small to yield stable data. It would be better to concentrate the sample in fewer grades to be sure of having stable data for those grades.

6. Sampling an adequate number of minority students. In most districts, school authorities will want to evaluate and compare the effects of desegregation on minority and on majority children. In districts where schools have only a small proportion of minority students, a larger number of classes will need

to be tested to obtain reliable data for minority students than would be required where the minority comprises a greater proportion of the student body.

Assuring that samples contain adequate numbers of minority students for purposes of analysis will be especially hard if these students are evenly distributed among classes. For example, if there are only one or two black students per class, a great many classes must be tested to obtain enough black students to yield stable data. When a district's evaluation budget is limited, such massive testing may not appear worthwhile, since testing a large number of intact classes will yield far more data on white students than would minimally be required for carrying out analyses.

However, alternative ways of securing data from an adequate number of minority students pose drawbacks of their own. For example, if minority students are called to a special room for testing, or are sent to take the test in a class other than their own, it will probably not be valid to compare the results of their tests with those of white students who are tested in their regular classrooms. Alternatively, if approximately equal numbers of blacks and whites are called to an auditorium, the novelty of being singled out for testing and taking the test(s) in a half-white, half-black setting casts doubt on test validity, especially the validity of racial

attitude data. Finally, if the problem of too few blacks per class is resolved by restricting testing to only those schools or classes where blacks comprise a larger minority, the generalizability of the data to the district as a whole is jeopardized.

Therefore, districts which have a small proportion of black students in some (or all) schools, and which also have limited evaluation funds, will need to consider reducing either the number of different kinds of tests given or the number of grades tested, in order to sample enough classes within selected grades to insure at least some stable data concerning the program's effects on minority students.

7. Accurate identification of student characteristics. In order to assess any differential effects which a desegregation program may have on majority and minority students, it is essential that the race or ethnicity of students be identified accurately. Whereas for sample selection purposes it is adequate to know merely the proportion of minority students in a school or class, for analytic purposes the racial or ethnic characteristics of each and every student tested must be accurately recorded. If SES is to be used as a variable in the evaluation design, it too must be accurately recorded for every student. It is usually a good idea to secure this information prior to the time of testing, especially when testing is scheduled for the end of the academic year. Provision must also be

made for accurately matching all the information identifying each child with each test he will take.

8. Stability of the desegregation program. Assessment of the cognitive and noncognitive effects of a desegregation program requires that the program not be changed in any significant way during the course of the planned evaluation, i.e., that no change take place in such important elements as the program objectives, the means of program implementation, etc. Since all the designs discussed below assume that the major independent and control variables and their levels are constant throughout the study, any change in these factors will invalidate the planned analysis of data.

In the event that there is a compelling reason for the district to make a major change in one or more program element(s) during the course of the study period, the evaluation design will need to be thoroughly reviewed and either altered or replaced by a new design.

9. Time of testing. Late spring is the preferred testing time for evaluating desegregation programs because by that time, students will have had maximum exposure to and experience in the desegregated environment. Special research situations may require districts to test at other times during the year (such as where a spring pretest was not feasible prior to the start of a program, and hence an early fall testing was substituted), but wherever possible--and especially when repeated testing is planned on a yearly basis--testing late in the academic year is to be preferred.

B. Designs for Desegregation Programs in Which Only Some Students are Program Participants

The effects of desegregation programs in which only some (rather than all) of a district's students are involved may be assessed by comparing the test scores of desegregated students (participants) with the test scores of nondesegregated students (nonparticipants). Differences in scores between the two groups are presumed to be due to the effects of desegregation. Such designs are based on the assumption that a suitable group of nonparticipants is available for test score comparisons with participants.

There are two main variants of such designs. In the simpler form, participants' scores are compared with nonparticipants' scores only after the desegregation program has been in operation for a specified period of time. In the more complex form, participants' scores are compared with those of nonparticipants both prior to desegregation and following desegregation.

In discussing these designs and in illustrating them schematically, time t_1 will refer to that test administration which takes place prior to the implementation of desegregation, and time t_2 will refer to that test administration which takes place following the implementation of desegregation.

1. Type I Designs: Postdesegregation comparisons only. In Type I designs, the desegregation program is evaluated by comparing the test scores of program participants and nonparticipants after the program has been in effect for a specified period of time. There is no testing of students prior to implementation of desegregation.

A Type I design is shown schematically in Figure V-1. The t_1 row indicates that no testing occurs prior to the implementation of desegregation; the t_2 row indicates that testing occurs for both groups following implementation of the program. The critical analytic comparison in this evaluation design is between the t_2 scores of participants and nonparticipants (comparison a in Fig. V-1).

a. Advantages of Type I designs. Type I designs have the advantage of requiring only a single test administration, thereby minimizing administrative costs of testing, scoring, etc. There is also no need to keep track of individual students over a period of time, thereby reducing clerical (record-keeping) requirements.

b. Disadvantages of Type I designs. The chief disadvantage of Type I designs is that the data which they yield are usually uninterpretable. In comparing only the t_2 scores of program participants and nonparticipants, any t_2 differences between the groups cannot unambiguously be interpreted

Time (t)	Students	
	Participants	Nonparticipants
Predeseg.: t_1	No Pretest	No Pretest
Program	Deseg.	No Deseg.
Postdeseg.: t_2	Posttest ←(a)→	Posttest

Fig. V-1. A Type I Design: Letter in ( ) indicates comparison referred to in text.

as being due to desegregation, since there is no adequate way of ruling out initial differences which may have existed between the two groups.

While the possibility of initial nonequivalence between program participants and nonparticipants is a common one in the evaluation of educational programs, it poses an especially serious problem in desegregation research because the children undergoing desegregation frequently are a distinctly nonrandom sample of all students in the district. When a desegregation program does not involve all the children in a school or school district, it is often the case that participants have volunteered for the program (or have at least agreed to cooperate with it). Whenever this situation occurs, the possibility exists that the attitudes and achievement of participating students may be different from those of students who have not volunteered or have not been willing to participate. Consequently, if differences are found between the test scores of participants and nonparticipants after the program has been in operation for a specified period of time, it is not legitimate to conclude that the differences are due to the effects of the program: except where assignment to the program has been completely random, there is always the possibility that the groups were not fully comparable at the start.

The fundamental problem in Type I designs of uncertainty concerning the initial comparability of participants and

nonparticipants is so serious (unless initial assignment was random) that these designs seldom yield definitive results. Consequently, Type I designs should be avoided.

However, in those instances where a design of Type I is the only design possible for a school district, a number of steps can be taken to increase the interpretability of the t_2 data. The scores of program participants should not be compared with a randomly selected group of nonparticipants, but rather should be compared with the scores of a group of nonparticipants matched to the program participants on as many important characteristics as possible.

The groups should be matched on the following personal variables: race, age, sex, previous academic achievement, and family background. Matching on environmental variables not affected by the desegregation program should also be attempted, e.g., matching on school or classroom balance characteristics which are not altered as part of the program. The matching may be based either on individual scores ("matched pairs") or on group means; however, the two methods require different types of statistical treatment.

Another good way of matching would be to use a comparison group composed of students who volunteered for the program but who could not be accommodated due to a limited number of openings. This procedure effectively matches participants and nonparticipants on the attitudes toward school,

desegregation, achievement, etc., which motivated students to volunteer to participate in the program.

2. Type II Designs: Predesegregation and postdesegregation comparisons. In Type II designs, program participants and nonparticipants are tested before the desegregation program begins and are tested after the program has been in operation for s specified period of time. Type II designs are illustrated in Figure V-2, where it can be seen that the spaces for participants and for nonparticipants are filled both at t_1 and at t_2.

A difference between program participants and nonparticipants in the degree to which test scores change from t_1 to t_2 (a versus b in Fig. V-2) or, alternatively, a difference in scores between program participants and nonparticipants at t_2 when there was no such difference at t_1 (c versus d in Fig. V-2) is interpreted as being due to the desegregation program.

a. Advantage of Type II designs. The advantage of Type II designs, when compared with Type I designs, is that it is possible to determine whether participants and nonparticipants were matched (i.e., had equivalent scores) on the dependent (effectiveness) measures prior to the inception of the desegregation program, and whether they were comparable on important personal and environmental characteristics. With such information, a determination can be made concerning

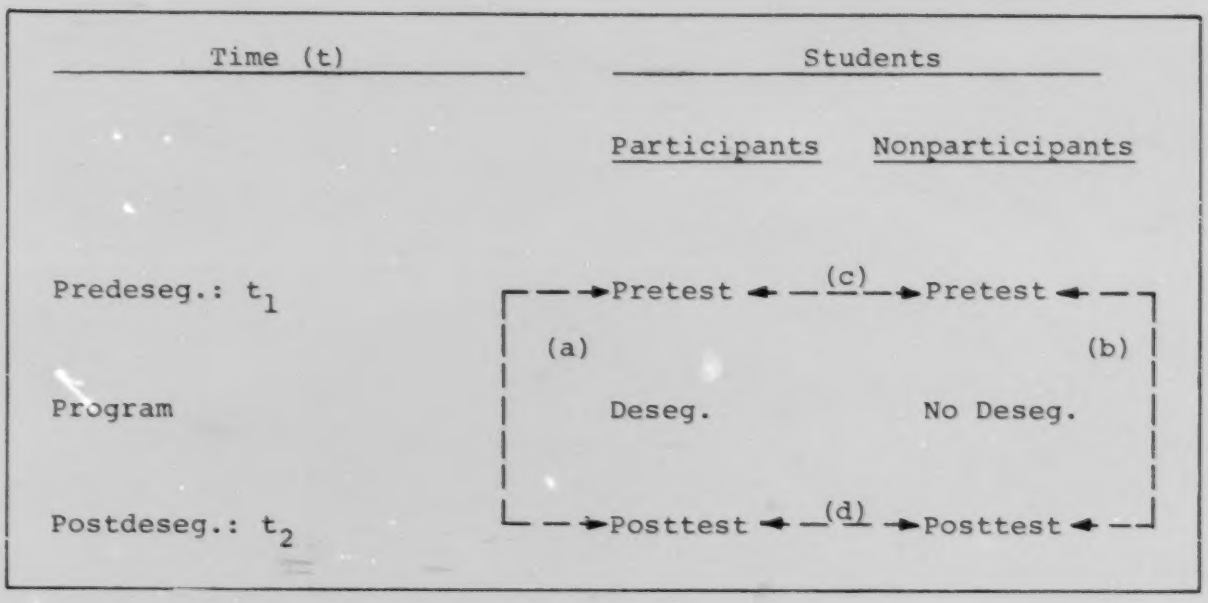


Fig. V-2. A Type II Design: Letters in ( ) indicate comparisons referred to in text.

whether or not the nonparticipants constitute an appropriate group with which to compare participants' t_2 scores. Therefore Type II designs result in data which are more interpretable than those obtained from Type I designs.

b. Disadvantages and problems of Type II designs. Type II designs are more expensive to complete than are Type I designs. The addition of a pretest doubles testing and scoring costs, and raises analytic (i.e., computer and interpretive) costs because the number of statistical comparisons which must be made is increased.

Type II designs also require greater administrative capability than Type I designs. First, districts must have their desegregation plans sufficiently detailed in advance to know which students will participate and which students will not. Second, districts must be sufficiently organized to complete the process of selecting and testing students prior to commencing the desegregation program, despite the fact that pressures for evaluation within the district are likely to be low at that time. Third, assignment of students to classes and schools must be made systematically, so that students will fall into the same levels of the independent and control variables for the posttest that they were in for the pretest. Fourth, the pretest sample must be selected with a feasible posttest schedule in mind, since students who are grouped together physically when the pretest is administered might otherwise be widely scattered when it is time for the

posttest. Finally, record keeping must be accurate. Pretest data for each student must be stored without error and in such a way that the t_1 scores can be readily retrieved and matched with the t_2 data.

Problems may be encountered regarding the equivalence of the t_1 and t_2 participant and nonparticipant groups. Districts must be alert to these problems since, if they occur and are not appropriately dealt with, the advantage of pretesting can be entirely lost.

The following problem might occur in some districts. The group selected to serve as nonparticipants may turn out to be an inadequate control group. For example, self-selection processes among students who volunteer to participate in the desegregation program may result in significant differences in the initial (t_1) attitude or achievement scores of participants and nonparticipants. If this nonequivalence is discovered early enough, additional sampling or a modification of the sampling plan might be feasible. Otherwise, it will be necessary to make use of analytic procedures which statistically "equate" the groups for their initial difference in scores.

Another problem may occur if students drop out of the program in a nonrandom pattern. This situation may result in the groups of participants tested at t_1 and t_2 being different from each other. For example, students who drop out of the program between t_1 and t_2 may be those who have the most

negative attitudes toward their school or their new teachers and classmates, or those whose academic performance is poorest. Thus students who remain in the program for testing at t_2 may represent a select subset of the original group. Since program participants and nonparticipants were matched for equivalence at t_1, and since it is assumed that nonparticipants are less likely to "drop out" than participants, the net result of selective dropping-out processes is apt to be nonequivalence of participants and nonparticipants at t_2. If this occurs, the advantage of a control group, obtained by matching participants and nonparticipants on the basis of the pretest, is lost.

Therefore, prior to the collection of t_2 test data, those students who have dropped out during the course of the evaluation study should be statistically eliminated from the t_1 sample, and the consequences of this deletion should be analyzed. If it is found that deletion of the dropouts destroys the initial (t_1) equivalence of participants and nonparticipants, it will be necessary statistically to eliminate some students from the nonparticipant group so as to restore the t_1 equivalence of participants and nonparticipants. If statistical equivalence cannot be restored, the district will have difficulty interpreting the t_2 data, and should seek the help of a consultant before proceeding.

Moreover, if analysis of the data reveals a systematic difference between program dropouts and non-dropouts with

respect to t_1 scores or with respect to personal or environmental characteristics, these facts should be noted, as they represent important findings in their own right.

Another potential source of difficulty in Type II designs lies in the fact that the reliability of a difference score is always lower than the reliability of the test itself. This means that any difference or change in score between t_1 and t_2 will be less reliable than either the t_1 or the t_2 test scores themselves. Because of the relative unreliability of individual change scores, districts should plan to employ statistical procedures which analyze the average differences between the scores of the groups of program participants and nonparticipants at t_1 and at t_2. Several methods exist for carrying out such analyses, and districts should obtain the assistance of a statistician to select an appropriate one. Extreme caution must be exercised if districts decide to compare changes in the individual scores of participants with changes in the individual scores of nonparticipants, especially if the sample size is small or if the reliability of the tests used is not very high.

Assuming that all of the methodological problems discussed above have been satisfactorily resolved, and that the design is therefore basically sound, some districts may nevertheless experience uncertainty in interpreting the results of their study if participation in the desegregation

program is confounded (coincides) with other educational changes.

In many districts, desegregation is accompanied by, and thus confounded with, other new or experimental programs which might themselves account for changes in scores on the effectiveness measures. For example, the introduction of study units on intercultural relations may affect racial attitudes; the provision of tutors, specialists, or reduced class size may affect achievement and attitudes; changes in the content of the academic curriculum (e.g., a new reading or math program) may result in test items on the posttest being better matched to curriculum content for program participants than for non-participants, or vice-versa. Under any of these conditions, posttest differences between the attitude or achievement scores of program participants and nonparticipants cannot be attributed with certainty to the change in racial composition.

In another common situation, desegregation is confounded with attendance at better quality schools. Desegregation programs frequently result in minority children from schools of low quality in poor neighborhoods being transferred to better schools in another neighborhood within their own school district or, in some cases, across district lines. The schools to which children are transferred may have better

facilities, more talented teachers, smaller classes, etc. than the schools previously attended by these students (in fact, the provision of "better quality education" is often the publicly stated reason for initiating the desegregation program). Under these conditions, it is possible that changes in the attitudes or achievement of desegregated students are due, at least in part, to the better quality education offered in the new school. In such a case, posttest score differences between program participants and nonparticipants cannot be attributed directly to the change in racial composition of the school attended.

As noted earlier in this chapter, because race and socioeconomic status (SES) are highly correlated in the United States, the desegregation process often results in a change in the social class characteristics of a child's classmates. Typically, desegregation results in an increase in the average socioeconomic status of the classmates of black students and a decrease in the average socioeconomic status of the classmates of white students. Since it has been demonstrated that the social class characteristics of a black student's classmates affect his academic achievement regardless of his own SES, it is possible that significant posttest differences in attitude or achievement scores between initially equivalent program participants and nonparticipants are due, at least in part, to the change in social class milieu. Thus

because of the confounding of race and SES, t_2 differences in attitude or achievement scores cannot, in most instances, be attributed specifically to the change in racial composition.

It must be stressed that the types of confounding discussed in the preceding three illustrations may actually be highly desirable from a pedagogical point of view, in that they may facilitate desired changes in student performance, e.g., more favorable attitudes, better achievement, etc. The fact that these sources of confounding are classified as "problems" is not meant in any way to suggest that schools should refrain from introducing other program, personnel, or facilities changes as part of, or in conjunction with, desegregation. Moreover, the discussion above certainly does not imply that districts should refrain from evaluating a desegregation program if that program is multifaceted. Carefully designed evaluation studies should be able to provide districts with at least some useful data on the overall effects of mixed desegregation programs.

However, confounds of the type described above do present interpretive problems because they limit the inferences which may be drawn from the studies in which they occur. The more that changes in racial composition are correlated with other changes which could affect students' scores on effectiveness measures, the more difficult it will be to attribute observed differences in test performance to changes in racial composition.

The ability to attribute program effects specifically to racial composition is valuable because it gives the district an empirical basis for either continuing or changing a desegregation program. From a logical point of view, if the program was planned to continue unchanged indefinitely, it would be of little practical importance to disentangle the effects of confounded program elements. But to the extent that the district may, in the course of time, wish to modify some element(s) of the program, it would be useful to know the specific effects contributed by the various factors, e.g., what effects are attributable to particular racial composition configurations?

3. Type IIA Designs: Predesegregation and postdesegregation comparisons with random student assignment. Type IIA designs are like Type II designs except that assignment of students as participants and nonparticipants is completely random. No self selection (volunteering) for the desegregation program is permitted. In principle, Type IIA designs are to be preferred over Type II designs because nonparticipants should represent a more adequate control group for participants. However, completely random assignment is seldom feasible in view of the variety of constraints (e.g., political, administrative, etc.) within which school districts operate. Moreover, while random assignment may reduce control group problems,

the other sources of confounding discussed in connection with Type II designs apply equally to designs of Type IIA.

C. Types of Designs When All Students are Program Participants

In the previous section of this chapter, the designs discussed (Types I and II) were based on the assumption that only some portion of the district's students were participating in the desegregation program. It was therefore possible to make use of a contemporaneous control group in those designs. However, in districts where the desegregation program includes all students (or at least all those in a given grade), a contemporaneous control group does not exist.

Since some form of control group is required for adequate assessment of program effectiveness, districts where all students are involved in the desegregation program must make use of designs with historical or longitudinal controls, i.e., designs employing predesegregation scores for control purposes. The predesegregation scores can be either the child's own scores prior to desegregation (Type III designs) or can be those of same-aged children in an earlier year, before the desegregation program was instituted (Type IV designs). Finally, it is possible to use both the child's own previous scores as well as the scores of other children in a single design (Type V designs), thereby incorporating the advantages of designs III and IV while eliminating some of the particular limitations associated with either design alone.

To keep the notation as consistent as possible with the notation used in the previous sections t_1 will refer to the

last testing session which precedes desegregation and t_2 will refer to the first testing session which follows desegregation. Where a design is expanded to include more than a single pre and posttest, the progressively earlier testings will be indicated by increasingly smaller subscripts of t (e.g., t_0, t_{-1},...t_{-n}), whereas the progressively later testings will be indicated by increasingly larger subscripts (e.g., t_3, t_4,... t_n).

1. Type III Designs: Longitudinal comparisons of scores for the same group of students. In Type III designs (Figure V-3) students' scores before desegregation are compared with their scores following desegregation. For example, scores of students in grade X before desegregation are compared with subsequent scores of the same students in grade X+1 following desegregation (comparison a in Fig. V-3). It is assumed that changes in test scores from pre to postdesegregation are attributable to the desegregation program.

a. Advantage of Type III designs. The advantage of Type III designs is that since exactly the same students participate in successive testings, there should be no question concerning the comparability of predesegregation and post-desegregation students on such important variables as race, sex, SES, and prior academic achievement, since each individual serves as his own control.

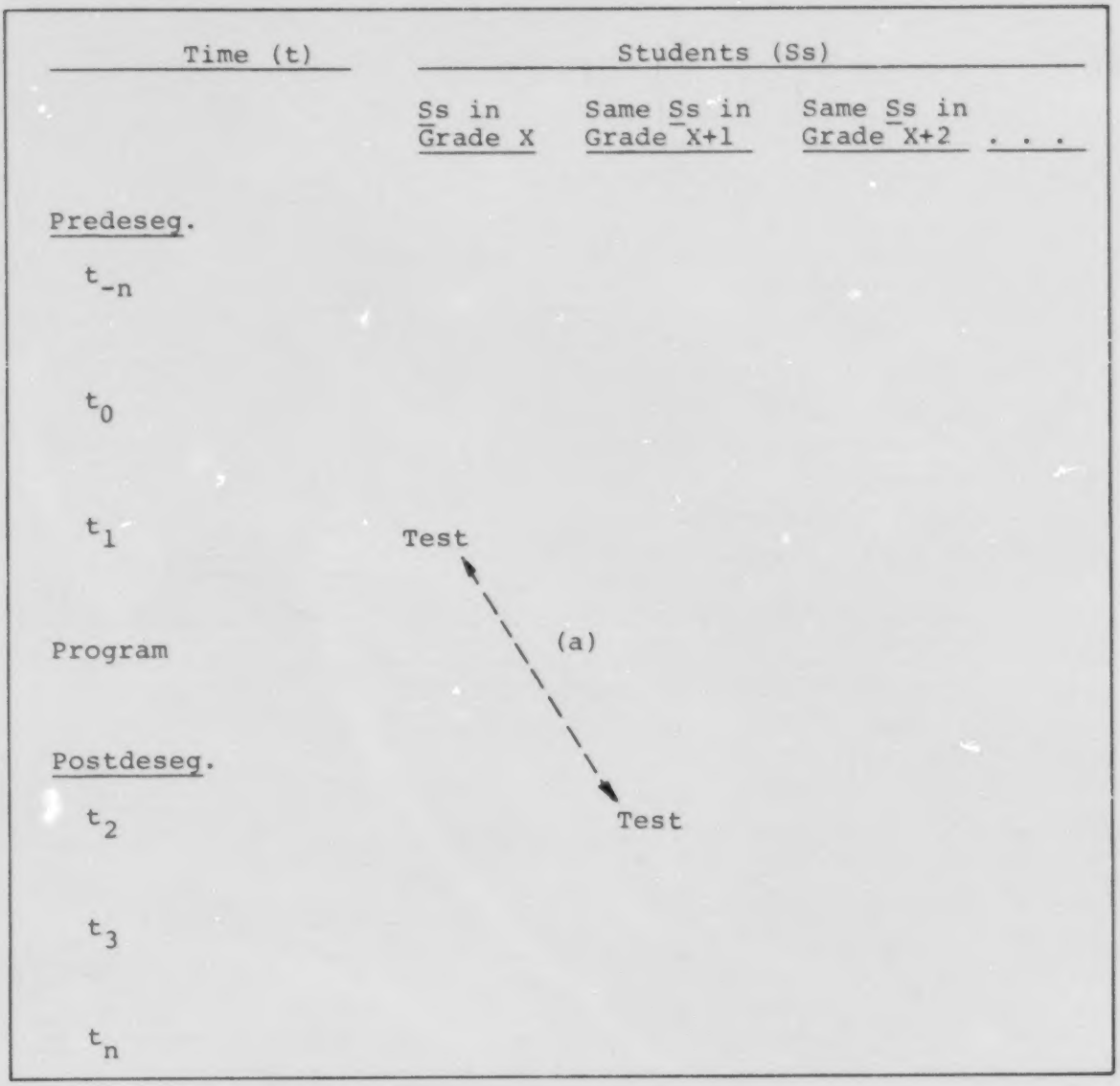


Fig. V-3. A Type III Design: Letter in ( ) indicates comparison referred to in text.

b. Disadvantages and problems of Type III designs. Designs of Type III share most of the administrative problems of Type II designs. If the same students are to be tested on various occasions and their scores compared, then assignments to classes and schools must be made in such a way that the environmental factors which are being used as control variables are held constant over time for each student. At a practical level, thought must be given to how best to sample and to assign students so that later retesting is feasible. Finally, since the same group of children is tested repeatedly, data must be stored in such a way that the previous scores of every student can be readily retrieved and accurately merged with new data.

A special administrative problem of Type III designs is the possibility of student attrition over the course of the study. In many districts the rate of student turnover during the period of the study will be high. Even if those who leave constitute a random sample of the entire student body (e.g., they have not left because their parents disapproved of the new desegregation policy), prior experience with longitudinal studies has shown that so large a proportion of the original sample may be lost over time that the remaining sample may be too small to obtain stable findings.

The longer the study continues, the greater the likelihood of student attrition. For this reason, districts

undertaking longitudinal studies which require repeated testing of the same group of students should plan to oversample (i.e., to test more students than will actually be needed for later analyses) in order to insure that there will be an adequate number of students remaining in the sample at the end of the study period. The oversampling should be proportional to the known rate of student turnover in the district and to the anticipated duration of the study, but at least a ten percent oversampling should be planned.

Another disadvantage of Type III designs is that they have limited interpretability. There is often no way to demonstrate that the observed test score changes would not have occurred without the desegregation program, since several factors which could be responsible for the changes in test scores from t_1 to t_2 are fully confounded with desegregation in Type III designs. Thus it is hard to determine whether changes in test scores are due to the desegregation program.

The most serious problem in interpreting data from Type III designs is that desegregation is confounded with maturation. Thus it may be argued that an observed change in score is maturational, i.e., that it is a change which normally accompanies development between certain ages, and one which would therefore have occurred even without desegregation. The longer the duration of the evaluation study, the more serious the problem of unknown maturational effects may become.

As the discussion in Chapter IV indicated, interpretation of scores on norm referenced achievement tests, on criterion referenced achievement tests, and on racial attitude tests requires baseline data from within the district (local norms or control groups) in order to properly evaluate changes in test scores following desegregation. Without such baseline data, observed changes in test scores cannot be compared with expected maturational changes in scores for district students over the time period studied. Since Type III designs do not include a contemporaneous control group and do not provide for comparisons with predesegregation baseline data for same-aged students, it is not possible to rule out maturation as the source of changes in scores on these tests.

Another difficulty in concluding with certainty that changes in test scores are due to desegregation is that a change in score may arise from familiarity with the tests. If the same tests are used at both t_1 and t_2, it is possible that at least some part of the change in scores is due to the children's prior experience in taking the tests. Since Type III designs include no students who have taken the tests but who have not been desegregated, changes in score associated with repeated test-taking cannot be assessed.

Following desegregation, a number of changes may take place within the schools and within the district which, on their own, could be responsible for any observed changes in students' attitude or achievement scores. The occurrence of these changes in the district makes it difficult to attribute changes in effectiveness test scores to the desegregation program.

For example, the change in test scores may be due to a change in the composition of the teaching faculty. When a school or school district undergoes desegregation, the pool of teachers may change. Teachers who have negative attitudes toward teaching minority students may leave and be replaced by teachers who are more motivated to work in an integrated setting and who, therefore, may tend to be more effective with racially mixed classes. Over a relatively short period of time, considerable transformation of the teaching faculty could occur via this process. Under such circumstances, it would be difficult to conclude that changes over time in attitude or achievement scores were directly due to the desegregation program. For this reason, careful attention should be paid to data on teacher turnover. In districts where such turnover has occurred to any marked extent, the change in teaching staff should be noted as an important outcome of desegregation in its own right, and should be cited as a possible factor in producing the observed changes in pupils' attitudes or achievement.

There are several other school factors, discussed above in connection with Type II designs, which often accompany desegregation and which could result in a predesegregation to postdesegregation change in the test scores of students. Briefly:

1. The general quality of schooling may improve as a result of the hiring of specialists, reduction of class sizes, new facilities, etc.;

2. Other new programs may be introduced along with desegregation, e.g., curriculum units designed to improve intercultural relations, new reading programs, etc.;

3. The SES composition of classrooms may change as a correlate of the change in racial composition.

Finally, in Type III designs it is difficult to rule out the possibility that the change in test scores may be due to influences external to the school. Since the schools are not insulated from the outside world, changes in student performance may reflect reaction to outside events rather than to school programs. For example, racial tension in the local community or elsewhere could affect students' racial attitudes or achievement; a teachers' strike during the year could affect achievement, etc. However, since all students are participants in the desegregation program and are subject to the same outside influences, the effects of desegregation and the effects of outside events are confounded, and are not easily separable.

2. Type IV Designs: Longitudinal comparisons of scores for different groups of students. In Type IV designs (Figure V-4) the test scores of students in a particular grade (or grades) are compared before and after desegregation. For example, scores of students in grade X before desegregation are compared with scores of other students in grade X following desegregation (comparison a in Fig. V-4). It is assumed that changes in the test scores of different waves of students from pre to postdesegregation may be attributed to the desegregation program.

a. Advantages of Type IV designs. Type IV designs are comparatively easy to administer since data for the same group of individuals do not need to be retained and matched with new data from later testings. Moreover, compared with Type II or Type III designs, less attention need be paid to the details of student assignments to particular classes or schools, either with regard to the impact of assignments on the environmental variables to be controlled, or with regard to the subsequent feasibility of testing students in new locations.

From an interpretive point of view, Type IV designs effectively enable the district to rule out maturational factors as a source of test score differences since, with grade (age) held constant, maturation can be assumed to be equivalent for students tested prior to desegregation and for students tested following desegregation. Familiarity with the test may

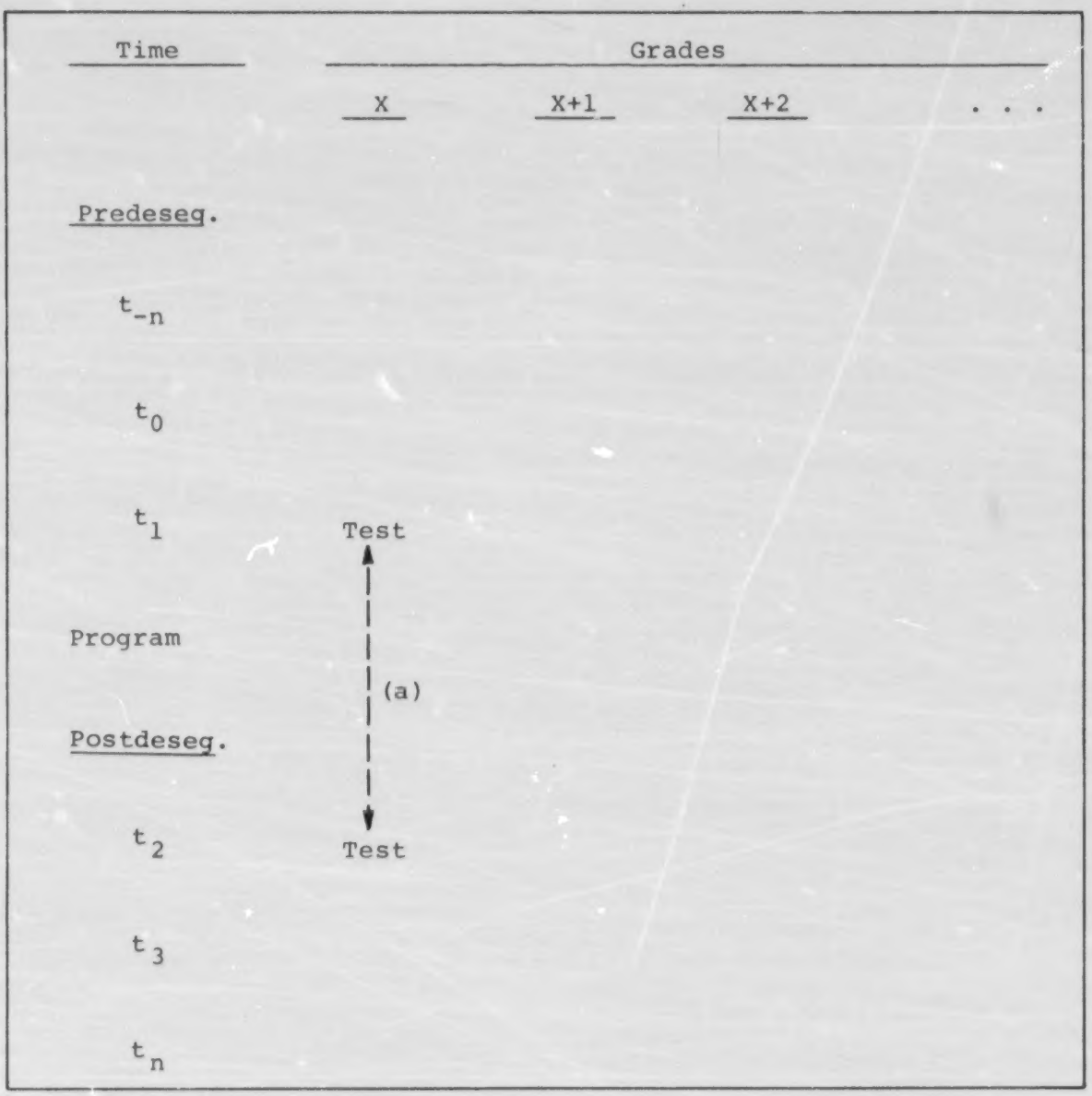


Fig. V-4. A Type IV Design: Letter in ( ) indicates comparison referred to in text.

also be eliminated as a possible source of confounding, since students of the same age should have taken the test an equal number of times and should, therefore, be equally familiar with it.

b. Disadvantages and problems of Type IV designs. With the exceptions just noted, Type IV designs share the interpretive difficulties associated with Type III designs. In short, it is difficult to attribute changes in test scores specifically to the desegregation program whenever:

1) The quality of schooling has changed concurrently with the implementation of the desegregation program, e.g., the school or district has introduced smaller classes, specialized staff and facilities, tutoring services, etc.;

2) Curriculum changes have occurred (e.g., new reading programs, intercultural relations units, etc.) which relate directly to performance on the effectiveness measures;

3) Teacher characteristics have changed following desegregation, as a result of nonrandom teacher turnover and replacement;

4) The SES composition of classrooms has changed as a correlate of changes in racial composition;

5) Community racial attitudes have changed as a consequence of contemporary social events.

In addition to these five sources of confounding, another potential problem for analyzing and interpreting data

in Type IV designs lies in the possibility that important characteristics of the school population may change during the course of the study. When an entire district is desegregated, and evaluation of the desegregation program continues over a period of several years, the school population may undergo systematic change during that time. For example, families opposed to desegregated schooling for their children may move away, and may be replaced by other families more willing to accept racially heterogeneous schools. In this case, the proportion of black and white students could remain stable, but the attitudes and/or achievement of students might change as a consequence of the change in parental attitudes. There is also the possibility that white families might leave and be replaced by black families, altering the proportion of black students over the time period studied and thus changing the levels of an important independent or control variable. Both examples illustrate population changes which would make it difficult to interpret changes in postdesegregation test scores.

While school district officials can do little to control turnover in the student body, provision should be made for checking the comparability of the population served (e.g., by SES, race, etc.) over the time period studied. If it is found that systematic changes in the student body have taken place or are in progress, the district should seek design and statistical assistance before undertaking any further testing

or data analysis.

3. Type V Designs. Type V designs (Fig. V-5) merge the essential features of Type III and Type IV designs. They combine repeated testing over time of the same group of students with the testing of successive waves of students of selected grade level(s). Preferably, testing for Type V designs should begin two years before the start of desegregation, in order to establish adequate baseline data for later comparisons.

With reference to Figure V-5, comparison of successive diagonals for the same grades (e.g., a versus b) may be used to assess differences between changes in scores during a year before desegregation and those during a year following desegregation. In this analysis, maturational effects are controlled. Comparison of diagonals between different grade levels (e.g., b versus d, with a and c as control data) may be used to assess the effects on students of being desegregated at different ages. The effects of desegregation on students of a particular age can be evaluated by analyzing vertical score changes within grades (e.g., e versus g). Moreover, comparisons between vertical changes in different grades (e.g., g versus h, with e and f as control data) can provide an additional way of evaluating the differential effects of desegregation as a function of age.

As more years ($t_3 \ldots . t_n$) are added to the basic design shown in Fig. V-5, any factors associated with the start of desegregation which might either inflate or depress scores

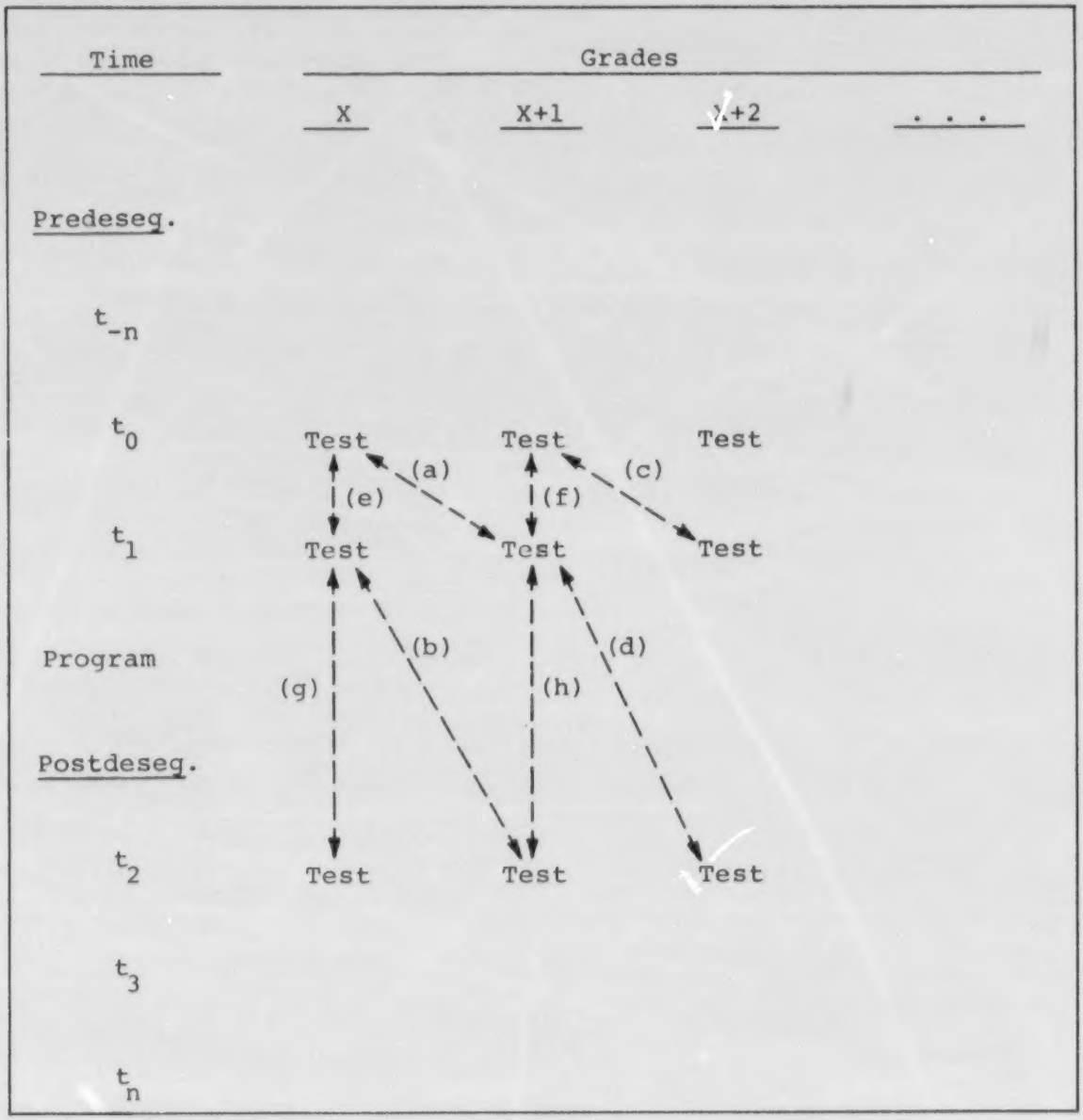


Fig. V-5. A Type V Design: Letters in ( ) indicate comparisons referred to in text.

temporarily (e.g., initial anxiety, excitement, confusion, etc.) would have time to dissipate, and thus more stable findings should be obtained. As more grades are added, groups of students can be followed for a longer period of time, and it becomes feasible to assess the long-term effects on students of having been desegregated at different times in their school careers.

a. Advantages of Type V designs. It is an advantage of Type V designs that longitudinal data for a particular group of students may be collected with maturational factors controlled. Moreover, depending on the care with which the sample is selected, it may also be possible to control familiarity with the test for at least some of the statistical comparisons.

The principal advantage of Type V designs is that they provide considerably more complete and more varied information than Type III or Type IV designs alone. Since at least two ways of analyzing the data (vertically for between-group information and diagonally for developmental information) are available, it becomes possible to validate one set of findings against the other. If all analyses result in relatively consistent findings, the strength of any conclusions which may be drawn is increased.

Several other types of interesting and important analyses are also possible when Type V designs are used, but

their description is not appropriate within the scope of this chapter in view of their complexity. Districts considering Type V designs should plan to work with a consulting statistician to define and examine all such possible analyses.

b. Disadvantages of Type V designs. Type V designs are the most costly of any of the designs discussed in this chapter because of the need for repeated testing of the same students over an extended period of time, because a larger sample size is usually required, because data collection must begin well in advance of desegregation, and because the greater complexity of analyses demands a higher degree of statistical expertise and analytic (e.g., computer) capacity.

Despite their greater cost and complexity, Type V designs may not provide wholly unambiguous results, since they are subject to many of the same sources of confounding noted above in connection with other designs, e.g., changes in school quality, changes in teacher characteristics, changes in student body characteristics, simultaneous introduction of multiple new programs, changes in SES characteristics of the educational environment, etc.

Another uncontrolled source of variation in test scores in Type V designs stems from the fact that all children born in the same year constitute a unique "cohort" of individuals who always share in common the experience of being a

particular age at a particular point in time.* Since no group before or after them ever experiences exactly the same set of events at the same ages as they, the cumulative life experience of any age cohort is somewhat unique.

In Figure V-5 each diagonal represents a different age cohort of students. While successive cohorts undoubtedly share many characteristics and hence are apt to be highly similar at any given grade level (e.g., as second graders, third graders, etc.), the uniqueness contributed by membership in a particular cohort may play a part in test data. Thus some part of the difference between the effects of desegregation at different ages (e.g., b versus d) may be due to membership in different cohorts, and may not simply be a function of age. The factor of cohort membership may also account for changes over time (e.g., e versus g) in the scores of successive waves of students at a given grade level, whether or not there has been a change in school programs.

Because Type V designs provide the greatest amount of information when carried out across several grade levels and when continued over a period of several years, they are

---

* This source of confounding also applies to Type III and Type IV designs, but has been reserved for discussion here in the interests of clarity of presentation.

best suited to districts which normally conduct an annual testing program in several or all grades. Moreover, since many districts do not formulate their desegregation plans more than one year ahead of time, the two year lead time required may be prohibitive except where districts already have prior information on all students. As a general guideline, however, districts lacking sufficient record-keeping, data storage, and data retrieval capability should probably not attempt Type V designs.

4. The special case of grade-a-year desegregation. Some school districts desegregate gradually, on a grade by grade basis, usually beginning during the first desegregation year with the first grade and desegregating one additional grade each year thereafter. When this pattern is followed, any child is either desegregated from his first grade in school or is not desegregated at all (at least not in elementary school). Accordingly when a grade-a-year plan is employed, some of the types of designs previously discussed cannot be used.

Neither Type II designs nor Type III designs are appropriate for evaluating grade-a-year desegregation plans. These designs require predesegregation and postdesegregation scores for each individual, while in grade-a-year plans individual students will have either pre or postdesegregation scores, but not both.

Type IV and Type V designs may be used, but only in those grades already reached by desegregation. For example, in the first year of desegregation, program effects could be assessed only for first graders; in the second year, they could be assessed for first and second graders, etc.* It would be six years before the effects of desegregation could be evaluated in the sixth grade. Hence, the higher the grade levels to be evaluated, the longer the study must continue.

Figure V-6 shows how a Type V design might look when applied to a grade-a-year desegregation plan. Development in segregated environments can be compared with development in desegregated environments by analyzing differences between successive diagonals for the same grades (e.g., a versus b). As additional grades are desegregated, comparisons between vertical changes in scores can be made (e.g., d versus f, with c and e as control data) to assess the cumulative effects at different ages of having been desegregated since the first year in school.

5. Summary Table. Major cost factors associated with each of the designs discussed in this chapter are summarized in Table V-1. Every type of design has features which could

---

* This restriction applies only to assessment of program effects on those children directly involved. "Spillover" effects on adjacent grades can be measured at any time.

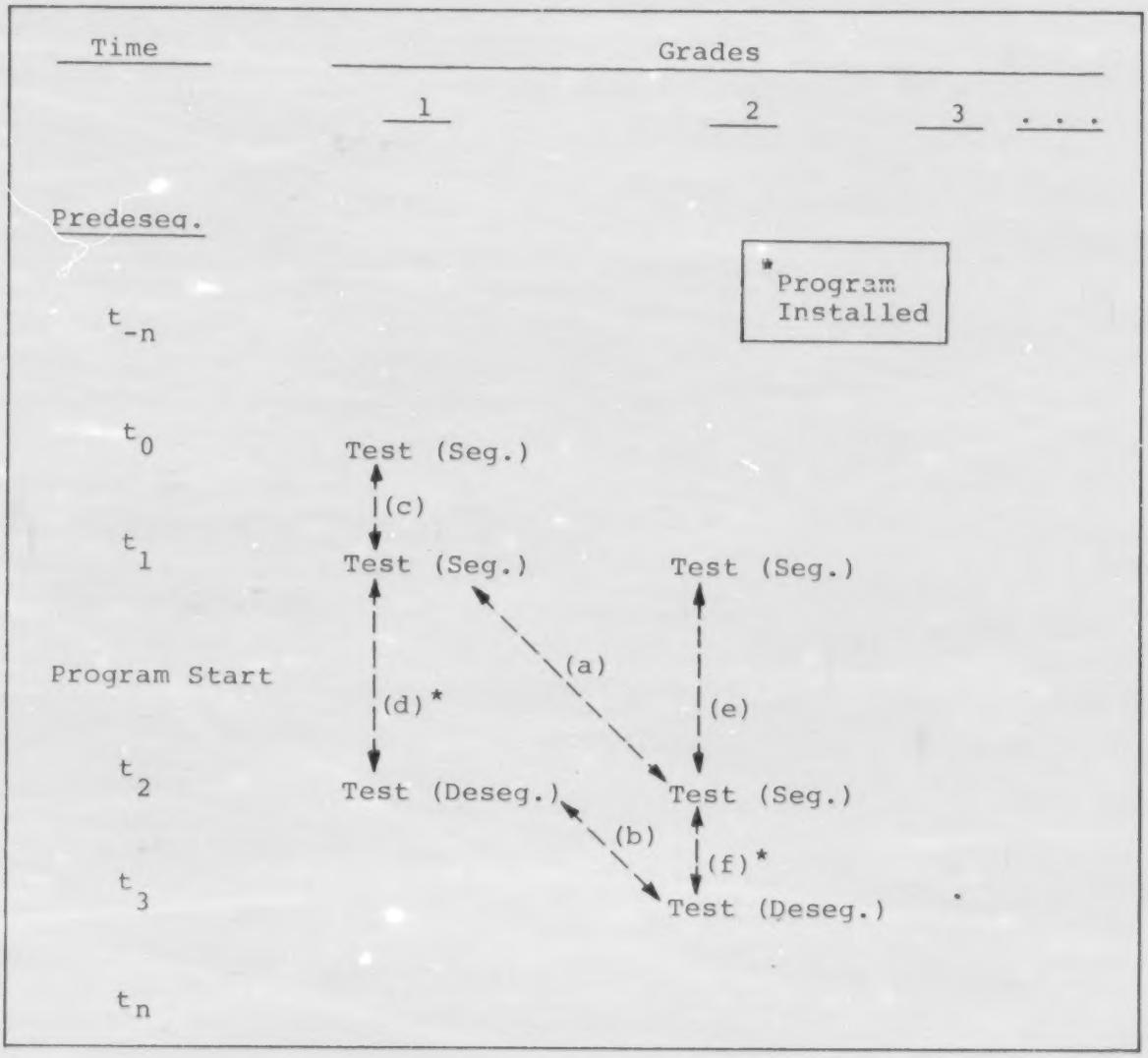


Fig. V-6. A Type V Design under Special Case of Grade-a-Year Desegregation: Letters in ( ) indicate comparisons referred to in text.

Table V-1

Summary of Cost Factors in Five Types of Evaluation Designs

Cost Factors	Designs				
	I	II	III	IV	V
Administrative Requirements					
Need to keep track of individual students	No	Yes	Yes	No	Yes
Length of time required for study	Short	Mod[a]	Mod	Mod	Long
Administrative complexity	Low	Mod	High	Mod	High
Data processing costs	Low	Mod	High	Mod	High
Need for statistical consultants	Low	Mod	High	Mod	High
Interpretive Limitations[b]					
Maturational confounds	No	No	Yes	No	No
Self-selection	Yes	Yes[c]	No	No	No
Loss of individual students	No	Yes	Yes	No	Yes
General change in student body	No	No	Yes	Yes	Yes
Test-taking experience	No	No	Yes	No	Maybe
Interpretable Data Yield	Low	Mod	Low	Mod	High

[a] Mod = Moderate

[b] Confounding factors which potentially limit the interpretability of all designs have been omitted from this table, e.g., other new programs simultaneous with desegregation, changes in school or teacher quality associated with desegregation, etc.

[c] For IIA: No

make it attractive to school districts as well as features which could make it unattractive. Each district must therefore decide which type of design will be most appropriate, given the nature of its own desegregation program, the characteristics of the student body and community it serves, and the fiscal resources and technical capability which are available for carrying out the evaluation.

Appendix

The Development of Racial Balance Indicators Based on the Probability of Occurrence of School and Classroom Distributions

The purpose of this appendix is to bridge the gap in Section III.B.3.b of these guidelines between the probability of occurrence of a distribution and the measure, or function, of this probability which can be taken as an indicator of racial balance.

The probability of occurrence of a particular distribution of students from minority group m to a set of schools or classrooms of specified capacity can be determined using the techniques of combinatorial analyses.*

Let:

N = the total number of students to be distributed;

N_m = the number of these students from minority group m;

$N_r = N - N_m$ = the remaining number of students;

C = the number of classes or schools to which the students are to be assigned;

n_c = the number of students in each class or school ($c = 1, 2, 3, \ldots, C$);

---

* W. Feller, op. cit.

n_{mc} = the number of students from minority group m in class or school c;

$n_{rc} = n_c - n_{mc}$ = the remaining number of students in class or school c.

It is assumed that N, N_m, N_r, C and n_c are all predetermined and fixed. Note that

$$\sum_{c=1}^{C} n_c = N, \quad \sum_{c=1}^{C} n_{mc} = N_m \text{ and } \sum_{c=1}^{C} n_{rc} = N_r.$$

Suppose that we begin by choosing students to fill the first classroom or school (c=1). A total of n_1 students must be chosen from the total number of students, N. The number of different ways in which we can choose n_1 students from the total population so that n_{m_1} are from minority group m and $n_{r_1} - (n_1 - n_{m_1})$ are not from minority group m is:

$$A_1(n_{m_1}, n_{r_1}) = \left[\frac{N_m!}{n_{m_1}!(N_m - n_{m_1})!}\right]\left[\frac{N_r!}{n_{r_1}!(N_r - n_{r_1})!}\right],$$

where, for example, $N_m!$ is read "N_m factorial," and is defined as

$$N_m! = (N_m)(N_m - 1)(N_m - 2) \cdot \cdot \cdot 3 \cdot 2 \cdot 1.$$

Each way of choosing n_{m_1} and n_{r_1} students from the total population constitutes a different arrangement. Two arrangements are different if one or more of the individual students chosen in one arrangement is not present in the other. However, each arrangement consists of the same number of m-students and r-students.

We now proceed to fill the second classroom or school by choosing n_{m_2} m-students and $n_{r_2} = n_2 - n_{m_2}$ r-students from the remaining population of $(N - n_{m_1} - n_{r_1})$ students. The number of different ways in which this can be done is:

$$A_2(n_{m_2}, n_{r_2}) = \left[\frac{(N_m - n_{m_1})!}{(n_{m_2})!(N_m - n_{m_1} - n_{m_2})!}\right] \times \left[\frac{(N_r - n_{r_1})!}{(n_{r_2})!(N_r - n_{r_1} - n_{r_2})!}\right]$$

This process is repeated until all students in the population have been assigned to classrooms or schools.

Therefore the number of different ways in which N students can be assigned to C classrooms or schools so that the classroom or school populations are $n_{m_1} + n_{r_1} = n_1$, $n_{m_2} + n_{r_2} = n_2$, . . . , $n_{mc} + n_{rC} = n_C$, is:

$$\prod_{i=1}^{C} A_i = A_1 \cdot A_2 \cdot A_3 \cdot \ldots \cdot A_C$$

$$= \frac{(N_m!)(N_r!)}{\left[\prod_{i=1}^{C}(n_{mi}!)\right]\left[\prod_{i=1}^{C}(n_{ri}!)\right]},$$

where Π denotes the continued product:

$$\prod_{i=1}^{C} x_i = x_1 \cdot x_2 \cdot x_3 \cdot x_4 \ldots x_{C-1} \cdot x_C ,$$

and where $n_1 + n_2 + \ldots + n_C = N = N_m + N_r$.

The number of ways in which N students can be assigned to C classrooms or schools so that a total of n_1 are in the first classroom or school, a total of n_2 are in the second, a total of n_3 are in the third, etc., <u>without regard to whether the students are from minority group m or not</u>, is

$$A_0 = \frac{N!}{\prod_{i=1}^{C}(n_i!)},$$

where, again, $n_1 + n_2 + \ldots n_C = N = N_m + N_r$.

Therefore the probability of occurrence of a particular distribution of $N = N_m + N_r$ students, for which the classroom or school populations are $n_{m1} + n_{m2} = n_1$, $n_{m2} + n_{r2} = n_2$, . . .,

$n_{mC} + n_{rC} = n_C$, can be found from:

$$P_D = \frac{\prod_{i=1}^{C} A_i}{A_0} = \frac{(N_m!)(N_r!)\prod_{i=1}^{C}(n_i!)}{(N!)\prod_{i=1}^{C}(n_{mi}!)\prod_{i=1}^{C}(n_{ri}!)} \qquad \text{(A-1)}$$

Here P_D denotes the probability of occurrence of distribution D. D is an arbitrarily assigned label (D = 1, 2, 3, . . .) used to identify different distributions. This formula is a generalized hypergeometric distribution, and has been used to calculate the probabilities for the example in Figure III-3. For that example, the formula reduces to:

$$P_D = \frac{(6!)^4}{(12!)(n_{m_1}!)(n_{m_2}!)(n_{r_1}!)(n_{r_2}!)} \; . \qquad \text{(A-2)}$$

The discussion in Section III.B points out that P_D is not a suitable racial balance indicator because its value is affected by factors such as school or district population size, and school district racial heterogeneity. In addition, the formula for P_D is rather unwieldy.

To construct a measure, or function, of P_D from which suitable racial balance indicators can be constructed, it is useful to work with several examples. One convenient example is the set of distributions appearing in Figure III-3. This

set (distributions 1 through 7) is repeated in Table A-1 together with various quantities which will be discussed subsequently in this appendix. Further examples of distributions which will be used to illustrate this discussion appear in Tables A-2 and A-3.

Finding a measure of P_D which is independent of N. The first step is to identify some function of P_D which is independent of N, the number of students being assigned. Table A-2 shows the least and most probable distributions of equal numbers of "minority" (m) and "majority" (r) students to two classrooms of equal size for three different values of N: N = 12 (distribution 1 and 4, transcribed from Table A-1); N = 32 (distributions 8 and 9); and N = 64 (distributions 10 and 11). Inspection of P_D for these six distributions indicates that:

- P_D corresponding to the least probable distribution (maximum imbalance) falls off very rapidly as N increases. For N = 64, $(P_D)_{min} = 4.88 \times 10^{-19}$;
- The values of $(P_D)_{min}$ become exceedingly small as N increases;
- P_D corresponding to the most probable distribution (maximum balance) also falls off with increasing N.

The precise way in which P_D depends on N is difficult to find. In many problems of this kind, it is useful to

work with the logarithm of P_D, rather than P_D itself, in an attempt to determine the dependence of P_D on N.

The logarithm (to the base 2) of P_D for each of the six distributions in Table A-2 is also shown in the table. Inspection of the values of $\log_2 P_D$ corresponding to the least probable distributions for the three values of N appearing in Table A-2 shows that they are all fairly close to N in magnitude. This pattern suggests that the quantity $(\log_2 P_D)/N$ (or rather, to get rid of the inconvenient minus sign, $(-\log_2 P_D)/N$) be examined to see how close it comes to having the required property.

Values of $(-\log_2 P_D)/N$ for the six distributions in Table A-2 are also entered in the table. Inspection of these values indicates that:

- The value of $(-\log_2 P_D)/N$ corresponding to $(P_D)_{min}$ tends towards one as N increases;
- The value of $(-\log_2 P_D)/N$ corresponding to $(P_D)_{max}$ tends towards zero as N increases.

Thus the quantity $(-\log_2 P_D)/N$ appears to show much less variation with N than P_D itself.

The last step is to write the expression for P_D (equation A-1) in a form which permits capitalizing on the observations made so far. To do this, a well-known approximation to the factorial (Stirling's formula) can be used:

$$x! \simeq \sqrt{2\pi x}\ x^x e^{-x}.$$

Replacing all the factorials in equation (A-1) by this formula, taking the logarithm, dividing by N, and multiplying by (-1), we obtain:

$$\frac{(-\log_2 P_D)}{N} \simeq T + (\text{other terms}) \qquad \text{(A-3)}$$

where:

$$T = \log_2 N - \frac{1}{N}\left[N_r \log_2 N_r + N_m \log_2 N_m + \sum_{i=1}^{C} n_i \log_2 n_i - \sum_{i=1}^{C} (n_{mi} \log_2 n_{mi} + n_{ri} \log_2 n_{ri})\right] \qquad \text{(A-4)}$$

The quantity T for the six distributions shown in Table A-2 appears in the last column of the table. The range of T is zero to one for each of the three values of N appearing in the table. T will be examined to see if it meets the requirements placed on acceptable measures of P_D.*

---

* The quantity T can also be obtained from information theory. Information theory leads to the development of uncertainty statistics, which represent an alternative, but less direct, way of describing student distributions. In the language of information theory, T is called the contingent uncertainty. It measures the extent to which uncertainty about a particular student's race is reduced by knowledge of the particular classroom or school to which the student is assigned. If schools or classrooms are maximally balanced, the uncertainty about a student's race is undiminished by knowledge of that

The quantity T can be shown to be independent of N for any distribution of students to schools or classrooms. Suppose we are dealing with a particular distribution, and suppose N is changed to γN (where γ is some constant) without changing the proportions of students from different groups in the population N, and without changing the number of schools or classrooms. (These conditions require the school or classroom capacities to change from n_i to γn_i.) Replacing N, N_r, N_m, n_i, n_{mi} and n_{ri} in equation A-4 with γ times each quantity, we obtain:

$$T(\gamma N, \gamma N_m, \gamma N_r, \ldots) = T(N, N_m, N_r, \ldots).$$

Table A-3 illustrates the independence of T on N. Three different values of N are considered. The racial heterogeneities are the same for each N (33.3%). For each value of N, T is calculated for three different distributions: an intermediate distribution (distributions 12, 15 and 18); the

---

student's school or classroom assignment, i.e., $T = T_{min} = 0$. If schools or classrooms are maximally unbalanced, the uncertainty about a student's race is greatly reduced by knowledge of that student's school or classroom assignment, i.e., $T = T_{max}$.

Discussions of information theory and its applications can be found in: F. Attneave, Application of Information Theory to Psychology, Holt-Dryden, New York, 1959; and W. R. Garner, Uncertainty and Structure as Psychological Concepts, John Wiley & Sons, Inc., New York, 1962.

least probable distribution (13, 16 and 19); and the most probable distribution (14, 17 and 20). Each intermediate distribution shows the same extent of racial balance, and the calculated value of T is the same for each (0.073) although the value of N is different for each. The values of T_{max} (corresponding to the least probable distribution for each N) are the same for each N (0.379), and the value of T_{min} (corresponding to the most probable distribution) for N = 54 is 0.000.

The values of T_{min} for N = 36 and N = 72 are almost zero (0.001), but not quite. The reason is that the class sizes do not permit minority students to be distributed in such a way as to reproduce in the classrooms the proportions of minority students in the populations of 36 and 72 students being assigned. The distributions closest to perfect balance in these cases are distributions 14 and 20. In both cases, the proportions of m-students are 0.35 in classroom 1 and 0.31 in classroom 2, compared to 0.33 in the student populations as a whole. Consequently, T_{min} is not exactly zero. In practice, such situations do not affect the interpretation of racial balance indicators when the number of possible distributions of minority students to classrooms or schools is sufficiently large.*

---

* Distributions involving small numbers of minority students are discussed in Chapter III.

In addition to being independent of N, the variation of T with P_D is also consistent with the operational definitions of racial balance. Racial balance is the extent to which the proportions of m-students in schools or classrooms approach the proportion of m-students in the student population being assigned. It can be shown formally* that T assumes its minimum value (zero, or very close to it as in the above examples) for the distribution corresponding to maximum racial balance, i.e., when the proportions of m-students in schools or classrooms are as close as possible to the proportion of m-students in the student population being assigned. Furthermore, the minimum value of T will be zero, or nearly zero, regardless of the racial heterogeneity of the student population (e.g., compare distributions 1, Table A-1, and 17, Table A-3). However, the maximum value of T (corresponding to maximum racial imbalance) does depend on the racial heterogeneity of the student population. This problem will be dealt with in the next part of this appendix.

Finally, although T is not equal to $(-\log_2 P_D)/N$, or to any other function of P_D which can be simply expressed, it appears to vary "monotonically" with P_D. That is, the ordering of values of P_D corresponding to different distributions (from higher to lower) is not affected by dropping

---

* Using Lagrange's method of undetermined multipliers.

the "other terms" in equation A-3. This monotonic property is all that we require of a measure of P_D in order to use it, in the manner described in Chapter III, to construct suitable racial balance indicators.

Eliminating the dependence of T on racial heterogeneity. The value of T_{max}, corresponding to maximum racial imbalance, is only equal to one if the racial heterogeneity of the student population is 50%.* If the heterogeneity is not 50%, T_{max} will be less than one. For example, in Table A-3, the heterogeneity is 33% and $T_{max} = 0.379$. Furthermore, T_{max} will be different for different heterogeneities.

This difficulty can be overcome by using a linear transformation of T, rather than T itself, as an indicator of racial balance. The particular linear transformation used in Chapter III is

$$B = 100 \times \left[\frac{T_{max} - T}{T_{max} - T_{min}} \right]$$

The quantity B is in the form used in Chapter III to define both school and classroom balance indicators. The indicator B has all the desirable properties of T which were discussed

* The magnitude of T, and hence T_{max}, also depends on the base of the logarithms used. The statement is true only for logarithms to the base 2. However, B is independent of the base chosen for the logarithms.

earlier in the appendix and, in addition, takes on the same value (i.e., zero) for the least probable distribution of a student population of any racial heterogeneity.

Table A-1

$N = 12;\ N_m = N_r = n_1 = n_2 = 6.$

D	n_{m_1}	n_{r_1}	n_{m_2}	n_{r_2}	P_D	$\log_2 P_D$	$(-\log_2 P_D)/N$	T
1	0	6	6	0	0.001	-9.966	0.831	1.000
2	1	5	5	1	0.039	-4.680	0.390	0.350
3	2	4	4	2	0.244	-2.035	0.170	0.082
4	3	3	3	3	0.433	-1.208	0.010	0.000
5	4	2	2	4	0.244	-2.035	0.170	0.082
6	5	1	1	5	0.039	-4.680	0.390	0.350
7	6	0	0	6	0.001	-9.966	0.831	1.000
					1.001			

Distribution of 6 m-students (from minority group m) and 6 r students to two classes, six students to each class

Table A-2

$N/2 = N_m = N_r = n_1 = n_2.$

D	N	n_{m_1}	n_{r_1}	n_{m_2}	n_{r_2}	P_D	$\log_2 P_D$	$(-\log_2 P_D)/N$	T
1	12	0	6	6	0	0.001	- 9.966	0.831	1
4	12	3	3	3	3	0.433	- 1.208	0.101	0
8	32	0	16	16	0	1.66×10^{-9}	-29.167	0.911	1
9	32	8	8	8	8	0.275	- 1.863	0.058	0
10	64	0	32	32	0	4.88×10^{-19}	-60.831	0.950	1
11	64	16	16	16	16	0.197	- 2.344	0.037	0

Most and least probable distributions of 12, 32 and 64 students to two classrooms of equal capacity. Racial heterogeneity is 50%.

Table A-3

$n_{m_1} + n_{m_2} = N_m$; $n_{r_1} + n_{r_2} = N_r$; $n_{m_1} + n_{r_1} = n_1$; $n_{m_2} + n_{r_2} = n_2$.

D	N	n_{m_1}	n_{r_1}	n_{m_2}	n_{r_2}	T	
12	36	4	16	8	8	0.073	
13	36	12	8	0	16	0.379	(max)
14	36	7	13	5	11	0.001	(min)
15	54	6	24	12	12	0.073	
16	54	18	12	0	24	0.379	(max)
17	54	10	20	8	16	0.000	(min)
18	72	8	32	16	16	0.073	
19	72	24	16	0	32	0.379	(max)
20	72	14	26	16	22	0.001	(min)

Most probable, least probable and intermediate distributions
of 36, 54 and 72 students to two classrooms
of unequal capacity.
Racial heterogeneity is 33%.